

MASTERARBEIT

B. Sc. Nadine Wappler

**Evaluierung der genomweiten
Genexpression bei
Rhodobacter sphaeroides
Stamm 2.4.1 im Kontext der
biologischen Wasserstofferzeugung**

Mittweida, 2014

MASTERARBEIT

**Evaluierung der genomweiten Genexpression bei
Rhodobacter sphaeroides Stamm 2.4.1 im
Kontext der biologischen Wasserstofferzeugung**

Autor:

B. Sc. Nadine Wappler

Studiengang:

Molekularbiologie/Bioinformatik

Seminargruppe:

MO12w1-M

Erstprüfer:

Prof. Dr. rer. nat. habil. Röbbbe Wünschiers

Zweitprüfer:

Prof. Dr. rer. nat. Petra Radehaus

weitere Betreuer:

Dipl. Ing. (FH) Lucy Stark

Einreichung:

Mittweida, 30. November 2014

Verteidigung/Bewertung:

Mittweida, 2014

Bibliographische Beschreibung:

Wappler, Nadine: Evaluierung der genomweiten Genexpression bei *Rhodobacter sphaeroides* Stamm 2.4.1 im Kontext der biologischen Wasserstoffherzeugung. - 2014. - 7 Seiten Verzeichnisse, 15 Seiten Inhalt in deutscher Sprache, 20 Seiten Paper in englischer Sprache. Mittweida, Hochschule Mittweida, Fakultät Mathematik, Naturwissenschaften, Informatik, Masterarbeit, 2014

Englischer Titel

Evaluation of the genome-wide gene expression in *Rhodobacter sphaeroides* strain 2.4.1 in context of biological hydrogen production

Kurzbeschreibung:

In dieser Arbeit werden Einflussfaktoren der Wasserstoffentstehung bei *Rhodobacter sphaeroides* Stamm 2.4.1 auf Genom- und Transkriptomebene untersucht. Zu diesem Zweck wurden eine Re-Sequenzierung sowie die Sequenzierung des Transkriptoms durchgeführt. Die Daten wurden anschließend bioinformatisch ausgewertet. Langfristiges Ziel ist die Optimierung der Fermentationsbedingungen von *Rhodobacter*. Die Ausbeute entstehenden Wasserstoffs soll maximiert werden, um das Gas als alternativen Energieträger verwenden zu können.

Die vorliegende Arbeit wurde zweisprachig erstellt und enthält zwei englisch sprachige Paper (Manuskripte).

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich während meines Studiums und der Anfertigung dieser Abschlussarbeit unterstützt haben.

Ein besonderer Dank geht an meinen Betreuer Herrn Prof. Dr. rer. nat. habil. Röbbbe Wünschiers für die Überlassung des interessanten Themas und die Unterstützung während der gesamten Entstehungszeit dieser Arbeit. Vielen Dank für das immer offene Ohr.

Ein großer Dank geht weiterhin an Frau Prof. Dr. rer. nat. Petra Radehaus für die Übernahme der Zweitbetreuung an der Hochschule Mittweida.

Weiterhin bedanke ich mich bei der gesamten AG Wünschiers. Hier ist besonders Lucy Stark zu nennen, welche mich im Labor angeleitet und beim Schreiben dieser Arbeit unterstützt hat. Eric Zuchantke danke ich für die Hilfe bei der bioinformatischen Auswertung.

Ich danke weiterhin der AG Jost Weber am Institut für Lebensmittel- und Bioverfahrenstechnik der Technischen Universität in Dresden für die Bereitstellung der Bakterienkulturen und die interessanten Diskussionen.

Nicht zuletzt möchte ich mich bei meiner Familie und meinem Lebensgefährten Marco für die Unterstützung und den Beistand während meines Studiums bedanken. Ohne sie wäre mir dieses Studium nicht möglich gewesen.

Liste der Paper

Diese Arbeit basiert auf den unten aufgeführten Papern:

- I **Wappler, Nadine; Zuchantke, Eric; Wünschiers, Röbbbe** (2014):
Re-Sequencing Rhodobacter sphaeroides strain 2.4.1 to facilitate
understanding of varying hydrogen evolution rates. Manuscript.

- II **Wappler, Nadine; Wünschiers, Röbbbe** (2014): Analyzing the
Transcriptome of Rhodobacter sphaeroides strain 2.4.1 in context
of biological hydrogen production. Manuscript.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	II
Abkürzungsverzeichnis	III
1 Einleitung	1
1.1 Biowasserstoffproduktion mit photosynthetischen Organismen	1
1.2 Sequenzanalyse	3
1.2.1 Sequenzierung von Genom und Transkriptom - Next-Generation Sequencing	4
1.2.2 Hilfsmittel für die Analyse von Sequenzdaten	5
1.3 <i>Rhodobacter sphaeroides</i>	7
2 Zielstellung	9
3 Manuskripte	10
3.1 Manuskript I	10
3.2 Manuskript II	11
4 Ausblick	12
5 Zusammenfassung	14
Literaturverzeichnis	IV
Selbstständigkeitserklärung	

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung der beteiligten Stoffwechselwege bei der Wasserstoffproduktion (Graphik abgewandelt aus Franchi et al., 2004).	2
Abbildung 2: Schematische Darstellung der RNA-Seq Methode (Wang et al., 2009).	5
Abbildung 3: Ausschnitt aus einem FastQ-File.	6
Abbildung 4: Mikroskopische Aufnahme von <i>Rhodobacter sphaeroides</i> (URL-1).	8
Abbildung 5: Schema des Prozesses der Wasserstoffgewinnung aus <i>Rhodobacter sphaeroides</i> mit Hilfe von Grünalgen.	12

Abkürzungsverzeichnis

BAM - *Binary Alignment/Map*

bp - Basenpaare (base pairs)

CBB - *Calvin-Benson-Bassham*

cDNA - komplementäre Desoxiribonukleinsäure (complementary desoxiribonucleic acid)

CO₂ - Kohlenstoffdioxid

DBMS - Datenbank Management System

DNA - Desoxyribonukleinsäure (*desoxiribonucleic acid*)

Gb – Gigabasen (giga bases)

GC - Gyanin/Cytosin

g - Gramm

H₂ - Wasserstoff

l - Liter

Mbp – Megabasenpaare (*mega base pairs*)

ml - Milliliter

mRNA - *messenger ribonucleic acid*

NGS - *Next-Generation Sequencing*

ORF - *open reading frame*

PHB - Poly-3-Hydroxybutyrat

RNA - Ribonukleinsäure (*ribonucleic acid*)

rRNA - *ribosomal ribonucleic acid*

SAM - *Sequence Alignment/Map*

SNP - *single nucleotide polymorphism*

SQL - *Structured Query Language*

µl - Mikroliter

µm - Mikrometer

1 Einleitung

Etwa 80% des derzeitigen weltweiten Energiebedarfs wird aus nicht erneuerbaren, fossilen Energieträgern gewonnen. Daraus resultiert neben der Klimaerwärmung auch eine enorme Emission von Schadstoffen. Wasserstoff (H_2) ist der Treibstoff der Zukunft (Franchi et al., 2004) und die sauberste Alternative zu fossilen Brennstoffen (Kim et al., 2014). Er zeichnet sich durch eine hohe Umwandlungseffizienz aus, ist recyclebar und umweltschonend. Wasserstoff ist kein primärer Energieträger sondern vielmehr ein Medium mit welchem Energie gespeichert und übertragen werden kann (Franchi et al., 2004). Mittels biologischer Gärung kann H_2 umweltfreundlich erzeugt werden. Dabei können verschiedene organische Produkte (Kim et al., 2014), wie beispielsweise Obst- und Gemüseabfälle oder Biomasse, (Franchi et al., 2014) als Ausgangsstoff genutzt werden. Die Gärung läuft bei Raumtemperatur und -druck ab, sodass kein zusätzlicher Energieaufwand notwendig ist (Kim et al., 2014).

H_2 wird nicht nur zur Energiegewinnung, sondern auch bei anderen Prozessen wie zum Beispiel bei der Refinierung von Fetten und Ölen oder dem Haber-Bosch-Verfahren eingesetzt. Neben den herkömmlichen Methoden der Wasserstoffproduktion wie beispielsweise der Dampfreformierung oder der thermischen Spaltung von Erdgas sowie der Elektrolyse kann Wasserstoff auch aus unterschiedlichsten organischen Abfällen gewonnen werden. Etwa 28% der Energie könnte beispielsweise in Form von Wasserstoff aus Saccharose (Rohrzucker) als Substrat zurückgewonnen werden. Die Produktion von biologischem Wasserstoff ist im Vergleich zu thermo- und elektrochemischen Prozessen weniger energieaufwändig (Das et al., 2001).

Die Wasserstoffgewinnung aus Biomasse und Bioabfällen ist eine CO_2 -neutrale Energiequelle (Franchi et al., 2004). Photosynthetische Bakterien wie *Synechocystis sp. PCC6803* (Montagud et al., 2013), *Clostridium butyricum* (Ferraz et al., 2014) oder *Rhodobacter sphaeroides* (Franchi et al., 2004) sind in der Lage, mit Hilfe von Sonnenlicht aus organischen Substraten Wasserstoff zu produzieren. Dieser Prozess wird als Photofermentation bezeichnet (Franchi et al., 2004).

1.1 Biowasserstoffproduktion mit photosynthetischen Organismen

Nicht-Schwefelbakterien sind die am besten untersuchten Wasserstoff produzierenden Bakterien. Die photosynthetischen Bakterien wachsen bei anaeroben Bedingungen und unter Lichtzufuhr photoheterotroph und verwerten organische Substrate. Dabei entsteht molekulares H_2 -Gas. Diese Organismen besitzen zahlreiche Alternativen für die Energiegewinnung, bei welchen kein Wasserstoff entsteht oder sogar verwertet wird. Die Kultivierungsbedingungen müssen daher angepasst werden. Biologischer Wasserstoff aus

photosynthetischen Organismen wird im Wesentlichen über zwei Stoffwechselwege produziert. Dies sind die Dunkel- und die Photofermentation. Bei der Photofermentation können theoretisch maximale Ausbeuten von Wasserstoff erzielt werden, wenn die Substrate komplett verwertet werden. Die Dunkelfermentation hat den Vorteil, dass organische Substrate sehr schnell zu Wasserstoff umgesetzt werden (Kim et al., 2014).

Bei photosynthetischen Bakterien wird der Gehalt an Wasserstoff, welcher anaerob aus organischen Substraten entsteht, durch die Interaktion diverser Stoffwechselwege bestimmt (wie in Abbildung 1 dargestellt). Die Wasserstoffentstehung wird dabei durch die Nitrogenase vermittelt und die Wasserstoffrückführung durch die Membran-gebundene uptake Hydrogenase katalysiert. Ein Überschuss an Kohlenstoff und Stickstoff in Kombination mit einer Limitierung von Licht führt zur Produktion des Speicherstoffs Poly-3-Hydroxybutyrat (PHB) (Franchi et al. 2004). Die PHB Synthase produziert Speicherstoffe in Abhängigkeit vom Substrat sowie dem Verhältnis von Kohlenstoff und Stickstoff (Yang et al, 2006).

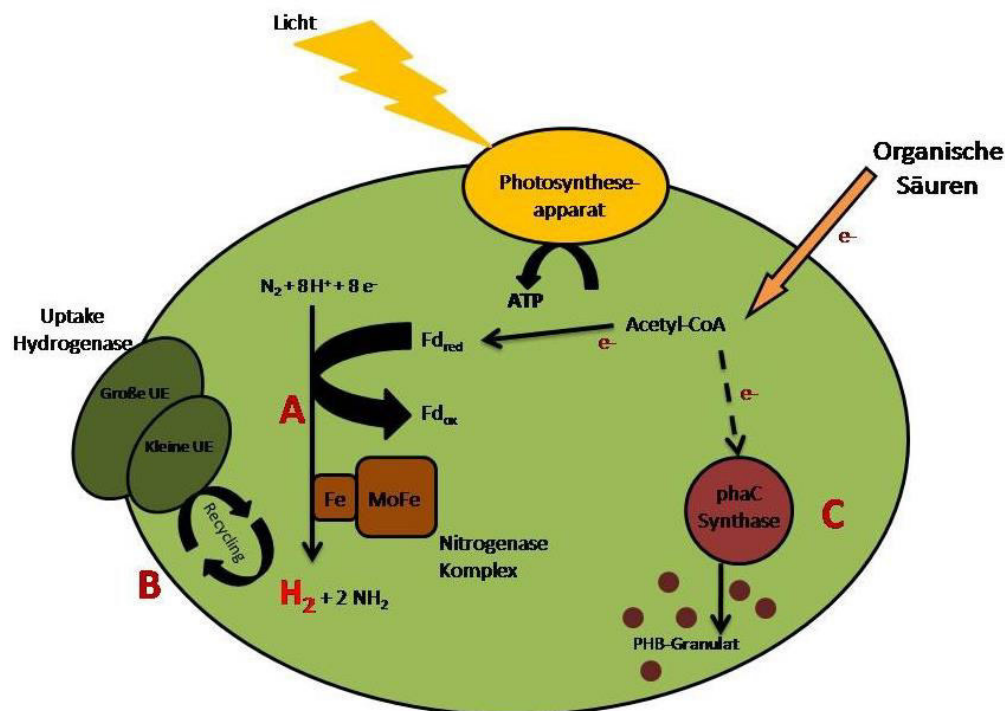


Abbildung 1: Schematische Darstellung der beteiligten Stoffwechselwege bei der Wasserstoffproduktion. A: Der Nitrogenasekomplex katalysiert die Wasserstoffentstehung. B: Die Membran-gebundene uptake-Hydrogenase sorgt für die Wasserstoffrückführung. C: Die PHB Synthase (phaC) ist in die Biosynthese des Poly-3-hydroxybutyrat-Granulats involviert (Graphik abgewandelt aus Franchi et al., 2004).

Nicht-Schwefelbakterien nutzen eine Vielzahl von organischen Säuren als Kohlenstoffquelle. Welcher katabolische Stoffwechselweg dabei genutzt wird, hängt von der Kohlenstoffquelle ab. Beispielsweise werden Acetat und Butyrat sehr leicht zu Acetyleinheiten konvertiert. Dabei entstehen mehr Polyhydroxy-Fettsäuren als Wasserstoff. Dies führt zu einer verminderten Wasserstoffproduktion. Werden Lactat und Succinat als Substrat eingesetzt

entsteht dagegen mehr H_2 im Vergleich mit Acetat und Butyrat. Lactat wird mittels der anaeroben Fermentation landwirtschaftlicher Abfälle gewonnen, sodass dies ein ideales Substrat für die photofermentative Wasserstoffproduktion darstellt (Kim et al., 2014). Franchi et al. (2004) beschrieben die Batch-Kultivierung von *Rhodobacter sphaeroides*, bei welcher Obst- und Gemüseabfälle als Substrat verwendet wurden. Hierbei wurden auch genetisch modifizierte Organismen eingesetzt, um die Nitrogenase-abhängige Wasserstoffphotoproduktion zu verbessern. Dies konnte erreicht werden, indem die Fähigkeit zur Rückführung von Wasserstoff (durch die Hydrogenase) oder der Speicherstoffbildung, abgeschaltet wurden. Dabei konnte gezeigt werden, dass die Doppelmutante eine beachtlich hohe Rate (im Vergleich zu Wildtyp und Einzelmutation) an Wasserstoffproduktion zeigte (Franchi et al. 2004).

1.2 Sequenzanalyse

Als Sequenzanalyse wird die Identifizierung der Abfolge von Nukleinsäuren bezeichnet. Mit diesen Informationen können Rückschlüsse auf Ähnlichkeiten zu orthologen und paralogenen Sequenzen getroffen werden. Aufgrund des hohen Datenaufkommens haben sich in der Bioinformatik zahlreiche Techniken für die Analyse von Sequenzen entwickelt. Dabei werden Sequenzen als Abfolge von Zeichen betrachtet, welche aus einer bestimmten Alphabet, also den Basen, resultieren. Neben der Ähnlichkeit zu anderen Sequenzen können auch Rückschlüsse auf Funktion, Abstammung und Verwandtschaft der untersuchten Sequenzen getroffen werden (Wink, 2011 S. 291).

1.2.1 Sequenzierung von Genom und Transkriptom - Next-Generation Sequencing

Die Next-Generation Sequencing (NGS) Technologie hat in den vergangenen Jahren einen bedeutenden Fortschritt bei der Untersuchung genetischer Informationen geliefert. Dazu zählen nahezu uneingeschränkte Einblicke in Genom, Transkriptom und Epigenom von Organismen. Das Prinzip des NGS basiert im Wesentlichen auf der Identifizierung der Basenabfolge kleiner DNA- oder RNA-Fragmente. Es ist möglich diesen Prozess auf Millionen von Reaktionen auszuweiten, welche parallel durchgeführt werden. Dieser Vorteil erlaubt die schnelle Sequenzierung ganzer Genome und Transkriptome. Neuste Geräte erlauben die Generierung hunderter von Gigabasen (Gb) an Daten in einem einzigen Sequenzierlauf. Im Jahr 2007 konnte ein einzelner Sequenzierlauf maximal 1 Gb an Daten liefern. Der enorme Fortschritt wird klar, wenn man betrachtet, dass diese Zahl vier Jahre bereits vertausendfacht werden konnte und ständig steigt. Dies ermöglicht Forschern Projekte schnell umzusetzen, da Datensätze in wenigen Stunden oder Tagen verfügbar sind.

Die **Sequenzierung eines Genoms** läuft im Wesentlichen in vier Schritten ab. Zunächst wird die genomische DNA extrahiert, in Librarys fragmentiert und parallel sequenziert. Danach werden die Reads mit einem Referenzgenome verglichen und im letzten Schritt eine Konsensussequenz aus den alignierten Reads erstellt (Illumina, Inc., 2013).

Das Illumina Genome Analyzer HiSeq/MiSeq System dient der automatisierten DNA-Sequenzierung durch Synthese. Die Library mit festen Adaptoren wird in Einzelstränge denaturiert und amplifiziert um Cluster zu bilden, welche klonale DNA-Fragmente beinhalten. Vor der Sequenzierung werden die Libraries mittels Linearisierungsenzymen in Einzelstränge gespleißt. Es folgt die Zugabe von fluoreszenzmarkierten Nukleotiden, welche komplementär an die Einzelstränge binden. Durch Analyse der Fluoreszenzsignale kann die genaue Basenabfolge bestimmt werden. Auch diese Technik wurde in den vergangenen Jahren stark verbessert. So können heute bis zu 600 Gb pro Lauf in wenigen Tagen sequenziert werden. Die Fehlerrate liegt bei etwa 2 %. Es können mehrere Proben gleichzeitig analysiert werden (Liu et al., 2012).

Für die **Analyse des Transkriptoms** in Organismen haben sich unterschiedliche Techniken etabliert. Dazu gehören hybridisierungs- und sequenzbasierte Ansätze. Neue Techniken erlauben neben der Quantifizierung auch das Mapping des Transkriptoms. Bei der RNA-Seq Technik (siehe Abbildung 2) wird die RNA in eine cDNA übersetzt und an einem oder beiden Enden mit Adaptoren ligiert. Jedes Transkript wird im Hochdurchsatzverfahren sequenziert. Die Fragmente sind üblicherweise 30-400 bp groß. Die Sequenzen können anschließend mit Referenzdaten verglichen oder de novo assembliert werden, was einen weiteren großen Vorteil dieser Sequenziermethode ausmacht – ein bestehender Modellorganismus wird nicht zwingend benötigt (Wang et al., 2009).

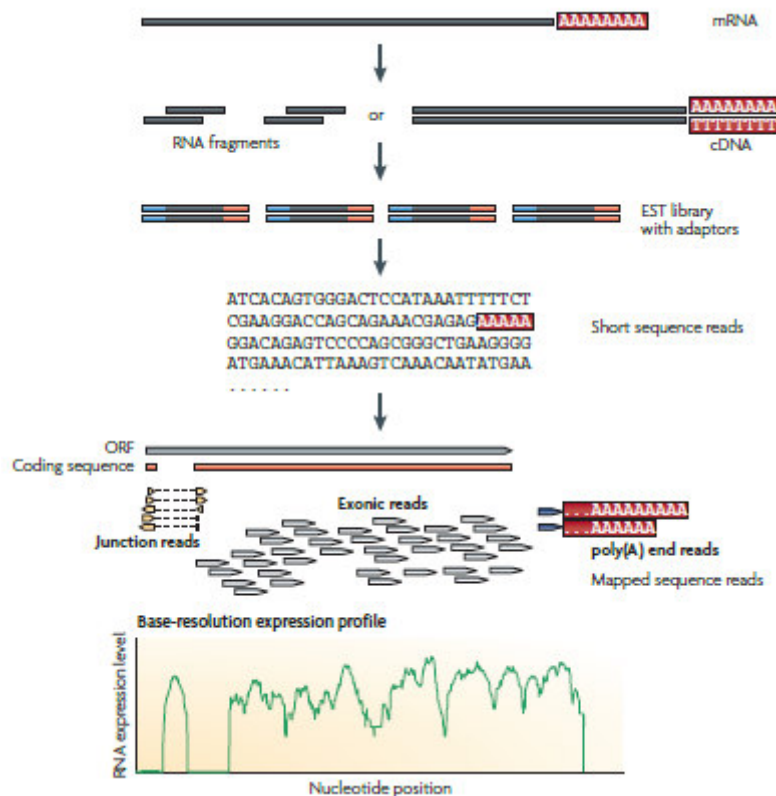


Abbildung 2: Schematische Darstellung der RNA-Seq Methode. Dabei wird die RNA fragmentiert oder in cDNA umgeschrieben. Jedes cDNA Fragment wird mit einem Adapter flankiert und die daraus resultierenden Sequenzen werden mit einem Referenzgenom verglichen und als *exonic-reads*, *junction reads* oder *poly(A) end reads* klassifiziert. Die drei Typen werden für die Generierung von Expressionsprofilen der Basenauflösung aller Gene benötigt. Die Sequenzierung erfolgt im Hochdurchsatzverfahren (Wang et al., 2009).

1.2.2 Hilfsmittel für die Analyse von Sequenzdaten

Für die Analyse von Sequenzdaten aus Hochdurchsatzmethoden werden unterschiedliche Werkzeuge aus der Bioinformatik eingesetzt, um Datenpipelines zu erstellen. Dadurch wird die sinnvolle Auswertung der gewaltigen Datenmengen möglich. Die in dieser Arbeit verwendeten Tools sind alle frei verfügbar und im Folgenden erläutert.

FastQC ist ein Tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) um rohe Sequenzdaten, beispielsweise aus Hochdurchsatzmethoden auf ihre Qualität zu untersuchen. Dies beinhaltet unterschiedliche Analysemethoden, um Probleme innerhalb der

Daten zu erkennen, bevor weitere Untersuchungen durchgeführt werden. Analysiert werden unter anderem die Sequenzqualität anhand des Phred-Scores, der GC-Gehalt oder das Vorhandensein von Sequenzduplikaten. Dabei können sowohl FastQ-Files, als auch SAM oder BAM Files (weitere Erläuterungen zu den Formaten befinden sich in diesem Abschnitt) analysiert werden (Andrews, 2014).

Um Sequenzdaten vor der eigentlichen Analyse zu bearbeiten, bietet das **FastX-Toolkit** (http://hannonlab.cshl.edu/fastx_toolkit/) eine Reihe von Kommandozeilen-Werkzeugen an. Dazu zählt der FastQ-Trimmer, welcher Sequenzen entsprechend des Phred-Scores verkürzt und Adapter entfernt. Der FastQ-Clipper entfernt Adapter und Linker aus der Sequenzierung (Schmieder et al., 2011).

Das FastQ-Format ermöglicht die Speicherung eines numerischen Qualitätsscores, Phred-Score genannt und der zugehörigen Nukleotidsequenz in einem File. Der Phred-Score beschreibt die erwartete Wahrscheinlichkeit, dass es sich bei der Sequenzierung ermittelten Base um einen Fehler handelt. Ein Beispiel für dieses Sequenz-Format, ist in Abbildung 3 dargestellt. Die erste Zeile beinhaltet den Titel und gegebenenfalls eine Beschreibung der der Sequenz, gefolgt von der Sequenzzeile. Die dritte Zeile besteht aus einem Identifier für den Qualitätsscore (+) und kann eine Wiederholung der Titelzeile beinhalten. Der Qualitätsscore befindet sich in der vierten Zeile (Cook et al., 2010).

```
@DE18INS655:23:C398RACXX:5:1101:1644:2034 1:N:0:ATCACG
NCCACTGCTTGATACGACACGGTTTCAGGTTCTTTTCACTCCCCTCGCCGGGGTTCCTTTTCGCTTTCCCTCACGGTACTGGTTCACTA
+
#1=DDFFF>FHFHHIGHIIIIIFHIIIIHIIIGIIIIIIIIIIIIIIIIIIHFBCEEEEDBDDDDDDDDDDDDDBDDDDDCDDCDDDD
```

Abbildung 3: Ausschnitt aus einem FastQ-File. In der ersten Zeile befindet sich der Identifier der Sequenz gefolgt von der Sequenz selbst, Identifier des Qualitätsscores und dem Qualitätsscore der Sequenz.

Die auf ihre Qualität untersuchten und mit dem FastX-Toolkit bearbeiteten Sequenzen müssen zum Zwecke der Auswertung auf eine Referenzsequenz gemappt werden. **Bowtie2** ist ein Alignmenttool, mit welchem die Referenzsequenz zunächst indexiert und anschließend die Sequenzdaten auf diese gemappt werden.

Die Ausgabe erfolgt im SAM Format, wodurch die Weiterverarbeitung mit anderen Tools möglich wird (Azam et al., 2014).

Das **SAM-Format** (Sequence Alignment/Map) dient der Speicherung alignierter Sequenzen. Die Alignments können anschließend mit der Software SAMtools weiter verarbeitet werden. Dazu zählen unter anderem die Indexierung und die Möglichkeit der Visualisierung von Alignments. Weiterhin ist mit SAM Tools die Umwandlung vom SAM- in das **BAM-Format** (Binary Alignment/Map) möglich. Dies ist ein Äquivalent zu SAM im binären Format und beinhaltet genau die gleichen Informationen (Li et al., 2009).

Die generierten BAM-Files können mit der Software **Artemis** weiter verarbeitet werden. Artemis ist ein Genom-Browser und Annotationstool. Weiterhin können Sequenzen verglichen und Bereiche, welche unterschiedlich oder identisch sind, markiert werden (Caver et al., 2010).

Zur weiteren Analyse der Expressionsdaten werden statistische Verfahren benötigt. **R** ist eine Softwareumgebung für statistische Berechnungen und Visualisierung (Wünschi, 2013). **DESeq** ist ein Paket innerhalb von R und gehört zum Bioconductor Projekt. Die Eingabe erfolgt in Form einer Tabelle, welche die zu verarbeitenden Daten enthält. Untersucht werden kann unter anderem die differentielle Expression von Genen (Anders et al., 2010).

MySQL ist ein Datenbank Management System (DBMS) , welches Structured Query Language (SQL), also eine standardisierte Sprache innerhalb des relationalen Datenbanksystems verwendet. Mit MySQL sind tabellenübergreifende Abfragen innerhalb der Datenbank möglich (Wünschi, 2013).

1.3 *Rhodobacter sphaeroides*

Rhodobacter sphaeroides wird der Alphaklasse der Proteobakterien zugeordnet (Kontur et al., 2012). Dieses Bakterium kommt ubiquitär im Boden und in Süßwasserlebensräumen vor (Porter et al., 2011). Der GC-Gehalt des 0,2-0,5 µm großen, begeißelten Stäbchenbakteriums ist mit 65-69 % vergleichsweise hoch. Das Bakterium zeichnet sich durch seine Vielseitigkeit an Stoffwechselwegen zur Energiegewinnung aus. Dazu gehören die aerobe Atmung bei hohem Sauerstoffpartialdruck und die anoxygene Photosynthese. Bei Licht und geringem Sauerstoffpartialdruck besitzt der Organismus die Fähigkeit zur Kohlenstoffdioxid/Stickstoff-Fixierung. Die anaerobe Atmung ist bei Abwesenheit von Licht und Sauerstoff durch Verwertung von Kohlenstoffäquivalenten und dem Vorhandensein eines Elektronenakzeptors möglich. Ist dieser ebenfalls nicht vorhanden, erfolgt die Energiegewinnung durch Gärung. Carotinoidmoleküle (Spheroiden und Spheroidon), welche in Bakteriochlorophyll sowie in carotinoid-bindende Komplexe eingelagert sind, sorgen für die charakteristische rote Färbung. Unter aeroben Bedingungen zeigt *Rhodobacter sphaeroides* nur eine leichte rosa Färbung. Bei semiaeroben Bedingungen zeigt sich eine dunkelrote Färbung aufgrund verstärkter Pigmentbildung. Die Regulation der Photosynthese ist sauerstoffabhängig (Metz, 2010).

R. sphaeroides besitzt nicht nur die Fähigkeit zur Bildung molekularen Wasserstoffs (Franchi et al., 2004), sondern ist auch bei der Entgiftung diverser Metalloxide beteiligt. Auch bei der Bioremediation spielt *Rhodobacter sphaeroides* eine Rolle (Giotto et al., 2006).

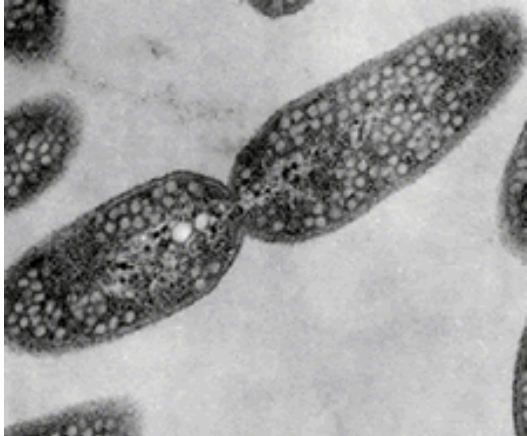


Abbildung 4: Mikroskopische Aufnahme von *Rhodospirillum rubrum* (URL-1).

Rhodospirillum rubrum besitzt zwei Chromosomen und fünf Plasmide (A-E) (Kontur et al., 2012). Chromosom I ist ~3.19 Mbp (NZ_AKVW01000001.1), Chromosom II ~0.94 Mbp (NZ_AKVW01000002.1), Plasmid A ~0,12 Mbp (NZ_AKVW01000003.1), Plasmid B ~0,11 Mbp (NZ_AKVW01000004.1), Plasmid C ~0,11 Mbp (NZ_AKVW01000005.1), Plasmid D ~0,10 Mbp (NZ_AKVW01000006.1) und Plasmid E ~0,04 Mbp (NC_009008.1) groß.

Es konnte bereits umfangreiches Wissen über Genregulation und Expression gesammelt und Genomanalysen über diesen Organismus durchgeführt werden. Aus dem Genomprojekt der *University of Texas* resultierte eine große Anzahl von DNA-Sequenzen für beide Chromosomen (Zhou et al., 2003).

2 Zielstellung

Das Ziel dieser Arbeit ist die Suche nach Parametern zum Zwecke der Prozessoptimierung bei der Kultivierung des photosynthetischen Nicht-Schwefelbakteriums *Rhodobacter sphaeroides* Stamm 2.4.1. Zu diesem Zweck wurde eine Re-Sequenzierung des Genoms (Manuskript 1) sowie die Sequenzierung des Transkriptoms (Manuskript 2) durchgeführt. Die Auswertung der Sequenzierungsergebnisse erfolgt im Kontext der biologischen Wasserstoffproduktion. Daraus sollen Rückschlüsse auf die Genexpression getroffen um zum einen schwankende Wasserstoffproduktionsraten besser verstehen und Prozessparameter anpassen zu können.

3 Manuskripte

3.1 Manuskript I

Re-Sequencing *Rhodobacter sphaeroides* strain 2.4.1 to facilitate understanding of varying hydrogen evolution rates

to be submitted to BMC Microbiology

Re-Sequencing *Rhodobacter sphaeroides* strain 2.4.1 to facilitate understanding of varying hydrogen evolution rates.

Nadine Wappler¹, Eric Zuchantke¹, Röbbbe Wünschiers¹

¹University of Applied Science Mittweida, Faculty MNI, Technikumplatz 17, 09648 Mittweida, Germany

E-Mail: nwappler@hs-mittweida.de

Abstract

Often, experimental results obtained with a certain bacterial strain differ due to strain evolution in a particular laboratory. We chose to perform a full genome re-sequencing of *Rhodobacter sphaeroides* strain 2.4.1 with a subsequent single nucleotide polymorphism (SNP) analysis, in order to identify deviations from the type strain. Our focus lays on an understanding of experimental variations of hydrogen production rates among strains from different laboratories. For an insight into changes of the genome sequence we focused our comparative analysis on SNPs at chromosome I and II as well as all five plasmids (A-E) with respect to strain 2.4.1. We found a total of 222 SNPs which caused 71 changes in protein coding genes.

1. Introduction

About 80% of the worldwide energy demand is covered by non-renewable fossil fuels. This results in emission of pollutants and global warming. Hydrogen (H₂), produced by renewable energy sources that often show daily or seasonal fluctuations, is an ideal future energy carrier [2]. Hydrogen is a clean energy source with a high energy density. It can be combusted or used as substrate in fuel cells to produce electricity. By biological fermentation, hydrogen can be produced with renewable substrates as starting material [4]. Some photosynthetic bacteria like *Rhodobacter sphaeroides* are capable of producing hydrogen from organic waste by photo-fermentation. *R. sphaeroides* belongs to the class of proteobacteria [6] and it is ubiquitous in soil [7]. It has various pathways for energy production, e.g. aerobic and anaerobic respiration or anoxygenic photosynthesis [8]. *Rhodobacter* has the ability for producing molecular hydrogen [9]. The enzymes nitrogenase, hydrogenase as well as enzymes of Calvin-Benson-Bassham pathway are involved in this process [5]. The strain 2.4.1 has two chromosomes and five plasmids (A-E) [6] and was sequenced for the first time in 2001 by the *University of Texas* [7]. Currently, five different strains, namely 2.4.1, WS8N, ATCC 17025, ATCC 17029 and KD131, are commonly used. They are characterized by a wide divergence between chromosome I and II. The size of chromosome I is relatively constant among all strains, while the size of chromosome II differs

[1]. For this study, we re-sequenced the genome of *Rhodobacter sphaeroides* strain 2.4.1 sub-strain TU Dresden. The aim is to understand metabolic variations from literature data observed with this sub-strain. We established a sequencing data processing pipeline for next generation sequence data.

The long term aim of this project is the optimization of *Rhodobacter* fermentation for hydrogen production. A theoretical approach is shown in figure 4. By co-cultivating green algae and *R. sphaeroides*, hydrogen can be produced efficiently from carbon dioxide and light.

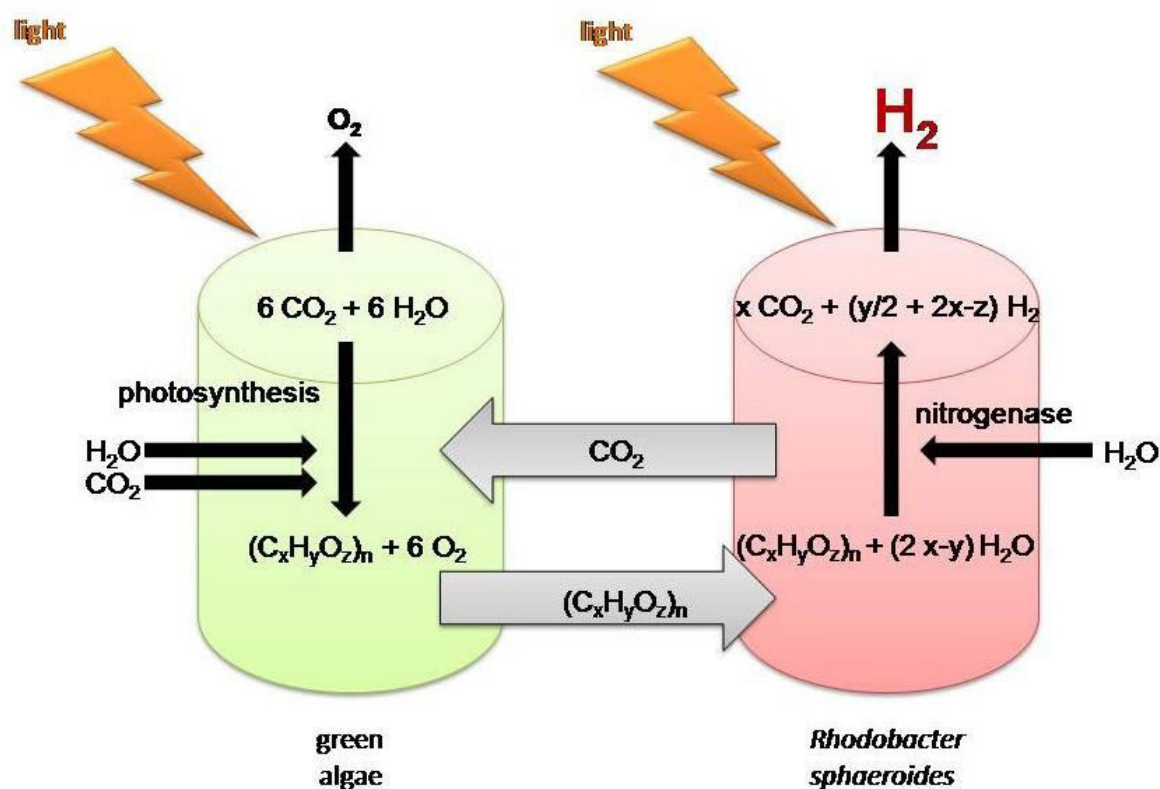


Figure 4: Schematic display of hydrogen production by *Rhodobacter sphaeroides* with the help of green algae. The bacteria uses carbohydrate equivalents from the metabolism of the algae. Hydrogen is derived from bacterial fermentation and carbon dioxide is fed back to the algae.

2. Material and Methods

2.1 Bacterial growth, harvest

Rhodobacter sphaeroides (DSM-No 158) was grown as batch culture in standard medium [3] as a pre-culture at 30°C and later as main-culture at 27°C, pH 7 in a 1 liter bioreactor. Vitamins and carbon sources [3] were added to the medium. Both cultures were illuminated by 600-Watt halogen lamps. The cells were harvested by centrifugation and frozen at -80°C until the next processing step.

2.2 Isolation of Nucleic Acids

DNA was extracted and isolated using the MasterPure™ Complete DNA and RNA Purification Kit (epicentre, Hessisch Oldendorf/Germany). Samples were thawed and placed on ice. 1 µl Proteinase K and 300 µl Lysis Solution Tissue & Cell were added. The cell pellets were dissolved by vortexing at full speed. Cell lysis was done at 65°C for 2 minutes and subsequently incubated for 1 minute at -20°C. 250 µl MPC Protein Precipitation Reagent were added and the mix centrifuged for 10 minutes at 14.100 rcf. The upper clear phase was transferred to a new tube with 500 µl ice-cold Iso-propanol, inverted 40 times and again centrifuged for 15 minutes at 14.100 rcf. Subsequently, the supernatant was removed and the pellet dried for 1 minute at 65 °C. The total nucleic acid pellet was resuspended in 50 µl TE-buffer for 15 minutes at 65 °C.

2.3 Digestion of RNA and sequencing

The isolated DNA had to be exempted from RNA contamination with RNase Cocktail™ Enzyme Mix (life technologies, Darmstadt/Germany). 2 µl were added to the tube followed by an incubation at 37°C for 30 minutes. The sequencing was performed with Illumina MiSeq (1x150 nt) by Eurofins Genomics GmbH (Ebersberg/Germany).

2.4 Data analysis

For genome data analysis (starting with fastq files), we solely used open source software. After a first assessment of quality by FastQC [10] we trimmed and clipped the sequence files using the FastX-Toolkit [11]. The indexing and mapping to the reference sequence from NCBI (NZ_AKVVW01000001.1) was performed with Bowtie2 [12]. SAMTOOLS [13] converted the incurred SAM- to a BAM-file, which was sorted and indexed afterwards. From the resulting Fasta-file a consensus sequence was created by BLASTing, which delivered a text file with SNPs. The resulting sequences were compared with the reference sequences from the NCBI Nucleotide database (<http://www.ncbi.nlm.nih.gov/nuccore/484336764>). A multiple genome sequence alignment of the different strains was done with Mauve [14].

3 Results and discussion

For a better understanding of evolutionary diversity we performed a multiple genome sequence alignment applying Mauve [13]. We used the five available genome sequences of *Rhodobacter sphaeroides* (ATCC 17025 [CP000661.1], ATCC 17029 [NC_009049.1, CP000579.1, CP000578.1], KD131 [CP001152.1, NC_011963.1, CP001153.1, CP001151.1], WS8N [NZ_CM001161.1, NZ_CM001164.1, NZ_CM001163.1, NZ_CM001162.1] and 2.4.1. [NZ_AKVVW01000001.1, NZ_AKVVW01000002.1, NZ_AKVVW01000003.1, NZ_AKVVW01000004.1, NZ_AKVVW01000005.1, NZ_AKVVW01000006.1, NC_009008.1]) from

the NCBI Nucleotide database and the resulting sequence from our re-sequencing (see figure 1).

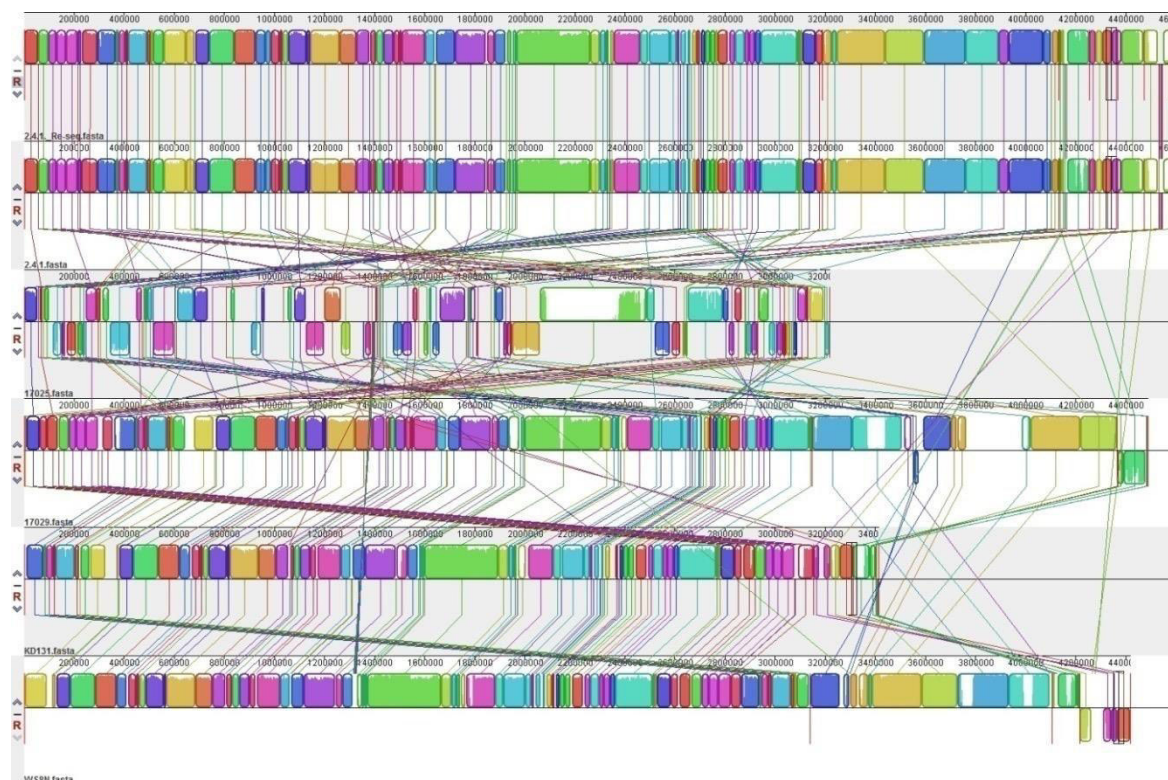


Figure 1: Multiple genome sequence alignment of six different strains of *Rhodobacter sphaeroides* with all available chromosomes and plasmids from the NCBI nucleotide database. Sequences from top to bottom: Re-Sequenced *Rhodobacter sphaeroides* 2.4.1, strain 2.4.1 (used as reference in evaluation of sequenced data), strain ATCC 17025, strain ATCC 17029, strain KD131 and strain WSN8.

Choudhary et al. (2007) postulated the divergence between the *Rhodobacter* strains of the both chromosomes as well as a varying size of chromosome II, which indicates a rapid evolution of this chromosomes caused by inclusion of genetic material, horizontal DNA transfer or rearrangements of genetic loci. They focused their analysis on strains 2.4.1, ATCC 17029 and ATCC 17025 and postulated a closer relationship of the former two. The variations can be explained by adaptations to different ecological niches, which results in formation of variable metabolism pathways [1]. For a better overview of evolutionary relationships between all five strains of *Rhodobacter sphaeroides*, we constructed a phylogenetic tree from Mauve output (see Figure 2) by using the package ape in R. In this study there was no evidence for genetic rearrangements compared with the reference strain 2.4.1. Based on the multiple genome sequence alignment and the resulting phylogenetic tree we observe a close relation between strains 2.4.1 and WSN8 and a strong divergence from strain ATCC 17025. This confirms the results from Choudhary et al. (2007).

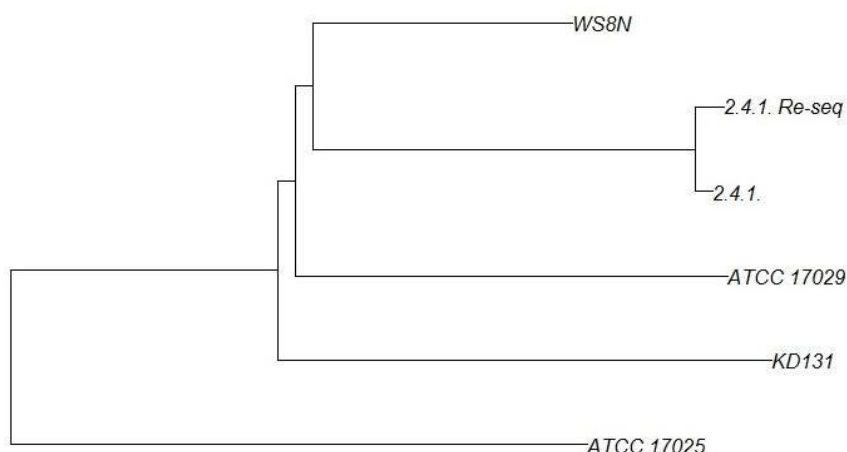


Figure 2: Phylogenetic tree of six different strains from *Rhodobacter sphaeroides* (re-sequenced *Rhodobacter sphaeroides* 2.4.1, strain 2.4.1, strain ATCC 17025, strain ATCC 17029, strain KD131 and strain WSN8). Tree is based on Mauve output by using package ape in R. Closely related genomes are on the same branches of the tree.

Our SNP analysis revealed a total of 222 SNPs, which caused 71 changes in protein coding genes in chromosome 1 as well as on plasmids D and E (see Table 1).

Table 1: Overview about identified SNPs.

	SNPs	genes	unchanged genes	SNPs in ORFs	SNPs in enzymes
chromosome I	128	6022	5974	48	18
chromosome II	0	1708	1708	0	0
plasmid A	5	217	217	0	0
plasmid B	0	209	209	0	0
plasmid C	2	177	177	0	0
plasmid D	83	200	188	12	2
plasmid E	4	68	67	1	0
total	222	8601	8540	71	20

Furthermore, we detected 15 SNPs within hypothetical proteins on chromosome 1, as well as 6 on plasmid D and one SNP in a hypothetical protein on plasmid E. 18 enzymes on chromosome I and two on plasmid D included a SNP. The results are shown in Table 2.

Table 2: Detailed information about found SNPs in enzymes and proteins.

	ID	product	SNP position		EC
			Start	Stop	
chromosome 1	RS0100005	16S rRNA	977	978	-
	RS0100455	uroporphyrinogen-III synthase	96436	96437	-
	RS0101090	tail tape measure protein	230577	230578	-
	RS0101140	phage capsid protein	238564	238565	-
	RS0101150	peptidase S49	240640	240641	-
	RS0101155	phage portal protein	241454	241455	-
	RS0101230	chromosome partitioning protein ParB	252370	252371	-
	RS0102200	pyruvate kinase	440842	440843	-
	RS0102425	inositol monophosphatase	490164	490165	-
	RS0102740	cold-shock protein	548631	548632	-
	RS0103000	cell envelope biogenesis protein AsmA	607146	607147	-
	RS0103215	DNA methyltransferase	646356	646357	-
	RS0103380	chromosome partitioning protein ParB	673356	673357	-
	RS0104305	AMP-binding protein	872185	872186	-
	RS0105170	ATP-dependent DNA ligase	1045084	1045085	6.5.1.1
	RS0105350	malto-oligosyltrehalose trehalohydrolase	1089251	1089252	-
	RS0105690	NADH dehydrogenase subunit D	1159765	1159766	1.6.5.3
	RS0106065	ATPase	1236258	1236259	-
	RS0106630	tRNA delta(2)-isopentenylpyrophosphate transferase	1353856	1353857	2.5.1.75
	RS0106805	sulfurtransferase	1388210	1388211	-
	RS0106985	exopolyphosphatase	1419395	1419396	-
	RS0107295	cobalt-precorrin-6x reductase	1488911	1488912	1.3.1.54
	RS0107355	membrane protein	1499471	1499472	-
	RS0108540	RNA polymerase subunit sigma-54	1774014	1774015	-
	RS0108600	flagellar hook protein FlgE	1790373	1790374	-
	RS0109650	radical SAM protein	2005967	2005968	1.97.1.4
	RS0109915	cellulose synthase	2063179	2063180	-
	RS0110470	hydrolase	2176196	2176197	-
	RS0110710	carbamoyltransferase	2229883	2229884	-
	RS0113155	MFS transporter	2726885	2726886	-
	RS0113445	membrane protein	2783990	2783991	-
	RS0114260	xylose operon repressor	2950652	2950653	-
	RS0114895	ATPase	3080432	3080433	-
plasmid D	RS0121740	TonB-dependent receptor	55425	55426	-
	RS0121750	diguanylate cyclase	58139	58140	-
	RS0121775	sulfate transporter	63800	63801	-
	RS0121780	universal stress protein UspA	65503	65504	-
	RS0121870	transposase	80558	80559	-
	RS0121925	sex-pilus assembly protein	89819	89820	-

Intergenic SNPs, which can be neutral or have an influence on gene-expression if they are located for example in a promotor region were found. Such SNPs are not further considered in this study. Genic SNPs can be neutral too, or lead to the substitution of amino acids that may affect enzyme activity (see figure 3).

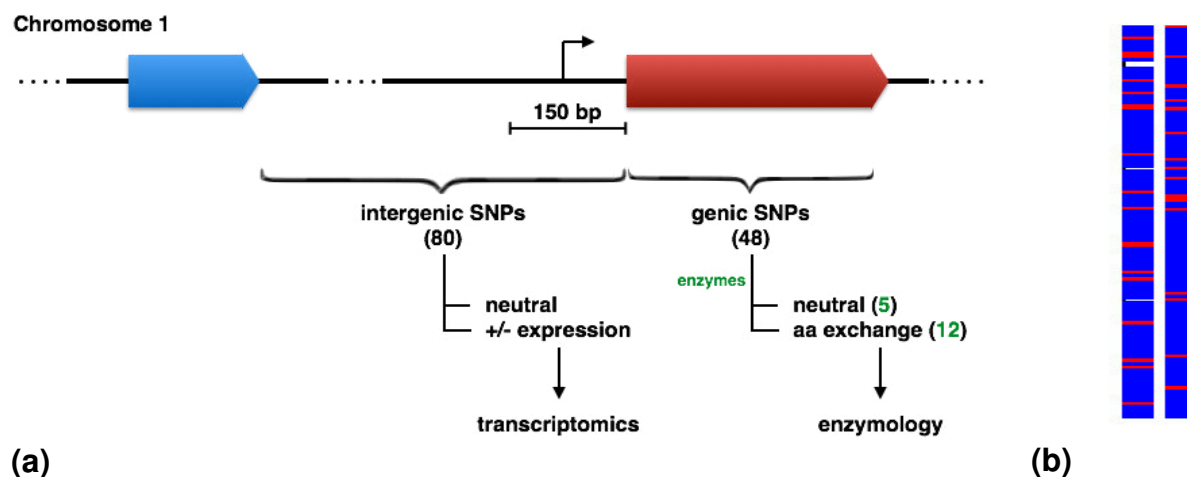


Figure 3: Schematic representation of effects caused by SNPs on chromosome I (a), which can have effects on transcriptomic regulation and enzymology. Such changes may lead to a new protein function. Found changes were mapped on chromosome I (b). White lines mark intergenic regions, blue lines regions with genes and red lines SNPs.

Due to their importance in metabolism, we searched for changes in enzymes, specifically those affecting base substitutions. 18 enzymes of chromosome I showed a total of 17 amino acid substitutions caused by a SNP, which led to changed protein sequences. These include the *uroporphyrinogen-III synthase* (P/L at position 28 of protein sequence), *peptidase S49* (T/I at position 266 of protein sequence), *pyruvate kinase* (P/R at position 100 of protein sequence), *DNA methyltransferase* (S/X at position 225 of protein sequence), *ATP-dependent DNA ligase* (R/C at position 46 of protein sequence), *malto-oligosyltrehalose trehalohydrolase* (T/A at position 326 of protein sequence), *NADH dehydrogenase subunit D* (A/E at position 2, M/T at position 3 and M/T at position 395 of protein sequence), *ATPase* [RS0106065] (S/T at position 317 of protein sequence), *exopolyphosphatase* (H/Y at position 260 of protein sequence), *RNA polymerase subunit sigma-54* (L/P at position 420 of protein sequence), *cellulose synthase* (H/R at position 249 of protein sequence), *hydrolase* (T/C at position 122 of protein sequence), *carbamoyltransferase* (M/L at position 1 and G/R at position 300 of protein sequence) and the *ATPase* [RS0114895] (T/P at position 599 of protein sequence). Unfortunately, none of the affected enzymes is directly involved in photo-hydrogen evolution. They may still have a pleiotropic effect. Furthermore, it cannot be ruled that SNPs located in hypothetical proteins, which are involved in regulatory processes, have an influence on hydrogen evolution rates. Analyzing the consequence of the found SNPs for

metabolism in general was not done in this study. By means of additional analysis of intergeneic SNPs it should be clarified if there are changes within promotor regions or structural implications of encoded proteins.

4 Conclusion

By re-sequencing the genome of *Rhodobacter sphaeroides* strain 2.4.1 we found a total of 222 SNPs. Thereof, 71 SNPs caused changes in protein coding genes. We focused further analysis on enzymes, which are only located on chromosome I. Within some enzymes, the SNP caused an amino acid substitution. None of the modified enzymes is directly involved in the process of hydrogen production. It can't be excluded that changes of other proteins have an indirectly influence on gas evolution. This should be clarified in further studies.

Acknowledgement

This work is performed in cooperation with the group of Jost Weber from the Technical University Dresden/Germany. For financial support we thank the State Ministry of Higher Education, Research and the Arts as well as the European Social Fund.

References

- [1] Choudhary, M.; Zanhua, Xie; Fu, Y. X.; Kaplan, S. (2007): Genome Analyses of Three Strains of *Rhodobacter sphaeroides*: Evidence of Rapid Evolution of Chromosome II. Journal of Bacteriology. p. 1914-1921.
- [2] Franchi, Elisabetta; Tosi, Claudio; Scolia, Giuseppe; Penna, Gino Della.; Rodriguez, Francesco; Pedroni, Paola Maria (2004): Metabolically Engineered *Rhodobacter sphaeroides* RV strains for Improved Biohydrogen Photoproduction Combined with Disposal of Food Wastes. Mar. Biotechnol. 6, 552-565, DOI: 10.1007/s10126-004-1007-y.
- [3] Hellwig, Rico (2014): Prozessoptimierung am Chemostatsystem zur kontinuierlichen photobiologischen Wasserstoffproduktion unter Verwendung von *Rhodobacter sphaeroides* DSM 158. Masterthesis, University of Applied Sciences Mittweida.
- [4] Kim, Dong-Hoon Kim; Lee, Ji-Hye; Kang, Seoktae Kang; Hallenbeck, Patrick C.; Kim, Eui-Jin; Lee, Jeong K.; Kim, Mi-Sun Kim (2014): Enhanced photo-fermentative H₂ production using *Rhodobacter sphaeroides* by ethanol addition and analysis of soluble microbial products. Biotechnology for Biofuels, 7:79

- [5] Kontur, Wayne S.; Ziegelhoffer, Eva C.; Spero, Melanie A.; Imam, Saheed; Noguera, Daniel R.; Donohue, Timothy J. (2011): Pathways Involved in Reductant Distribution during Photobiological H₂ Production by *Rhodobacter sphaeroides*. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 2011, p. 7425–7429
- [6] Kontur, Wayne S.; Schackwitz, Wendy S.; Ivanova, Natalia; Martin, Joel; LaButti, Kurt; Deshpande, Shweta; Tice, Hope N.; Pennacchio, Christa; Sodergren, Erica; Weinstock, George M.; Noguera, Daniel R.; Donohue, Timothy J. (2012): Revised Sequence and Annotation of the *Rhodobacter sphaeroides* 2.4.1 Genome. J Bacteriol., 194(24): 7016–7017.
- [7] Mackenzie, Chris; Choudhary, Madhusudan ; Larimer, Frank W.; Predki, Paul F.; Stilwagen, Stephanie; Armitage, Judith P.; Barber, Robert D.; Donohue, Timothy J.; Hosler, Jonathan P.; Newman, Jack E.; Shapleigh, James P.; Sockett, R. Elizabeth; Kaplan, Jill Zeilstra-Ryalls & Samuel (2001): The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. Photosyn. Res. 70: 19-41.
- [8] Metz, Sebastian (2010): Blaulichabhängige Genregulation in *Rhodobacter sphaeroides*. Untersuchungen zur physiologischen Funktion der Blaulichphotorezeptoren AppA, CryB und LOV. Inaugural - Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften. Justus - Liebig - Universität Gießen.
- [9] Porter, Steven L.; Wilkinson, David A.; Byles, Elaine D.; Wadhams, George H.; Taylor, Stephen; Saunders, Nigel J.; Armitage, Judith P. (2011): Genome Sequence of *Rhodobacter sphaeroides* Strain WS8N. J Bacteriol., 193(15): 4027–4028.

Used software and version:

- [10] FastQC, Version 0.11.2
- [11] FastX-Toolkit, Version 0.0.13
- [12] Bowtie2, Version 2.0.0-beta6
- [13] Samtools, Version 1.0
- [14] Mauve, Version 2.3.1

3.2 Manuskript II

**Analyzing the Transcriptome of *Rhodobacter sphaeroides* strain
2.4.1 in context of biological hydrogen production**

to be submitted to BMC Microbiology

Analyzing the Transcriptome of *Rhodobacter sphaeroides* strain 2.4.1 in the context of biological hydrogen production

Nadine Wappler¹, Röbbbe Wünschiers¹

¹University of Applied Science Mittweida, Faculty MNI, Technikumplatz 17, 09648 Mittweida, Germany

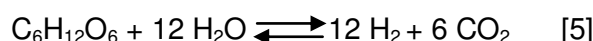
E-Mail: nwappler@hs-mittweida.de

Abstract

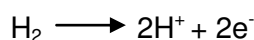
In order to achieve our long term objective of optimizing *Rhodobacter sphaeroides* (strain 2.4.1) fermentation parameters for hydrogen (H₂) evolution, we analyzed the transcriptome of this strain. In the future, this kind of energy recovery could be realized by an environmentally friendly co-cultivation of *Rhodobacter sphaeroides* and a green algae. For understanding experimental variations of H₂ production rates between a sub-strain of *Rhodobacter sphaeroides* (strain 2.4.1) cultivated at the TU Dresden and other strains, we determined transcripts that affect either the enzymes *nitrogenase*, the *hydrogenase* or the *Calvin-Benson-Bassham* (CBB) cycle [5]. The experiment was carried out with several glutamic acid concentrations as a nitrogen source. The goal was to judge the influence of this substrate on hydrogen production.

1 Introduction

Hydrogen gas is considered to be an efficient future energy carrier. It can be converted to electricity by fuel cells. With this kind of energy recovery there is not accompanied by the formation of harmful by-products. Several microorganisms have the ability to produce hydrogen gas. Photofermentative cyanobacteria, like *Rhodobacter sphaeroides* are able to produce hydrogen [6] during anaerobic, photoheterotrophic growth [5]. The amount of H₂ produced is depended on the substrate and mediated by various metabolic pathways [1]. The gas evolves under nitrogen deficient conditions by the *nitrogenase* enzyme complex using sunlight [7] and organic acids [6] as energy source:



It should be mentioned that the gas is a by-product of nitrogen fixation. Next to the *nitrogenase*, the *hydrogenase* and enzymes of CBB pathway are involved in hydrogen metabolism [5]. The enzyme *hydrogenase* has the function to recover hydrogen gas whereas members of CBB cycle reduce carbon dioxide and build cell mass [1]. *Hydrogenases catalyse the following reaction:*



There were many efforts in terms of optimizing hydrogen production rates. For example, Kim *et al.* (2014) postulated an increase of hydrogen evolution by the addition of ethanol, which suppresses the production of soluble microbial products [4]. Furthermore, immobilization on a solid matrix enhances the gas rate up to the range of liters per hour and measured per liter of medium. If not immobilized, *Rhodobacter* usually produces around 80 - 100 ml molecular hydrogen per hour and liter of medium [6]. The targeted knockdown of genes responsible for hydrogen uptake or formation of storage substances led to considerably higher hydrogen evolution rates as compared to the wild-type strain [1]. In this study, we sequenced the transcriptomes of *Rhodobacter sphaeroides* strain 2.4.1 applying different carbon sources. The aim was to determine transcripts involved in hydrogen production for facilitate understanding varying evolution rates of hydrogen. Prior to this study, a re-sequencing of the genome of *Rhodobacter* had been performed. Some single nucleotide polymorphisms could be identified but there was no evidence those are responsible for varying hydrogen production rates.

The long term aim of the project is the optimization of *Rhodobacter sphaeroides* fermentation parameters for hydrogen production. A theoretical approach provides the co-cultivation of a green algae and *R. sphaeroides*. By this process hydrogen can be produced efficiently from carbon dioxide and light by photosynthesis of the algae and *nitrogenase* of the bacteria.

2 Material and Methods

2.1 Bacterial growth, harvest and preparation with RNAlater

Rhodobacter sphaeroides (DSM-No 158) was grown as batch culture under anaerobic, photosynthetic conditions in standard medium [3] as a pre-culture at 30°C and later as main-culture at 27°C, pH 7 in a 1 liter bioreactor. Vitamins and carbon sources [3] were added to the medium, which differed only in the concentration of glutamic acid (0.1 g/l, 1.5 g/l and 5.0 g/l). All cultures were illuminated by 600-Watt halogen lamps. The cells were harvested by centrifugation, resuspended in 1:5 RNAlater (life technologies, Darmstadt/Germany) and frozen at -80°C until the next processing step.

2.2 Preparation of instruments and reagents

For sterilization of instruments autoclaved DEPC-treated (0.05% (v/v)) water and RNase AWAY (Carl Roth GmbH & Co. KG, Karlsruhe/Germany) were used. Solutions and reagents were used only for work with RNA.

2.3 Isolation of total RNA

RNA was extracted and isolated using MasterPure™ Complete DNA and RNA Purification Kit (epicentre, Hessisch Oldendorf). Samples were thawed and placed on ice. After

centrifugation by 5.000 rcf for 10 minutes the supernatant containing medium and RNAlater was removed. 1 µl Proteinase K and 300 µl Lysis Solution Tissue & Cell were added. The tubes were vortexed until the pellet was dissolved. Lysis was carried out at 65°C for 2 minutes followed by cooling at -20°C for 1 minute at -20°C. 250 µl MPC Protein Precipitation Reagent were added and centrifuged for 10 minutes at 14.100 rcf. The upper clear phase, containing nucleic acids, was transferred to a new tube with 500 µl ice-cold Iso-propanol, inverted 40 times and centrifuged again for 15 minutes at 14.100 rcf. Subsequently, the supernatant was removed and pellet was dried for 1 minute at 65 °C. 50 µl of TE buffer were added and the samples were resuspended for 15 minutes at 65 °C.

2.4 Digestion of DNA, depletion of rRNA and sequencing

DNA was degraded by TURBO DNA-free™ Kit (life technologies, Darmstadt) following the standard protocol. Depletion of ribosomal RNA was done by mRNA-ONLY™ Prokaryotic mRNA Isolation Kit (epicentre, Hessisch Oldendorf/germany), using the standard protocol as well. The sequencing was performed with Illumina MiSeq (1x150 nt) by Eurofins Genomics GmbH (Ebersberg/Germany).

2.5 Data analysis

For transcriptome data analysis (starting with fastq files), we solely used open source software. After a first assessment of quality by FastQC [8], the data was trimmed and clipped using the FastX-Toolkit [9]. Indexing and mapping to the reference sequence from NCBI (NZ_AKVW01000001.1) was performed with Bowtie 2 [10]. SAMTOOLS [11] was applied to convert the incurred SAM- to BAM-files, which were sorted and indexed afterwards. In the next step, a read count analysis was carried out by Artemis [12]. After normalization of the data by the package DESeq (R) [13], we analyzed the expression values in terms of different expression and genes of interest. Differentially expressed genes were plotted onto a Boehringer Biochemical Pathway map [14].

3 Results and discussion

First of all we considered the mapping rate of all three samples by using the mean alignment rate, shown in Figure 1. The alignment rate indicates the percentage of the sequence data could be mapped on the reference sequence. Surprisingly, there was no evidence for gene expression for all five plasmids, whereas for both chromosomes a good mapping rate of at least 85 % was achieved. Hence, we excluded the plasmid sequences from further analysis.

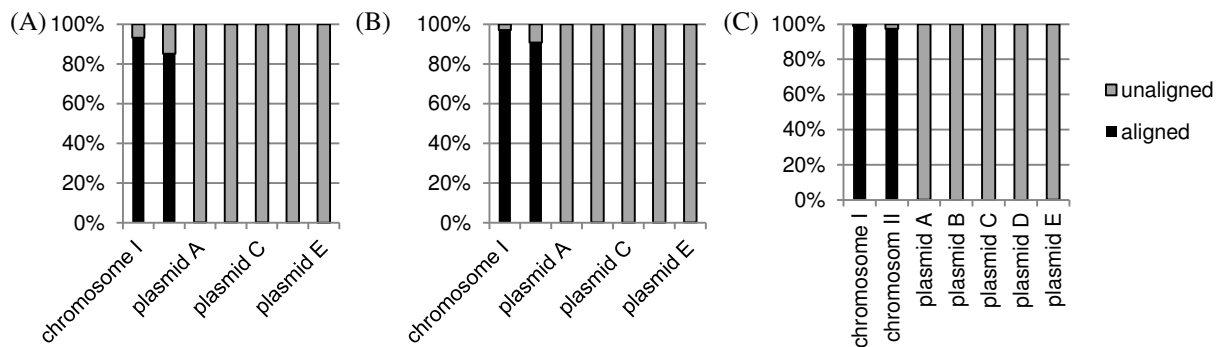


Figure 1: Overview about aligned und unaligned sequences from the mean of three triplicates: 0,1 g/l (A), 1,5 g/l (B) and 5,0 g/l (C) glutamic acid in medium. Black bars show aligned and grey bars unaligned regions.

A overview of all expressed transcripts from chromosome I and II is shown in Figure 2 in the form of two heat maps. Transcripts of all three samples were compared to each other, were different colors represent different expression levels. Data were clustered by similar expression pattern. It follows that such transcripts are located in analog regions in the heat map. As shown in Figure 2, transcripts are highly differentially expressed in chromosome I (A) as well in chromosome II (B).

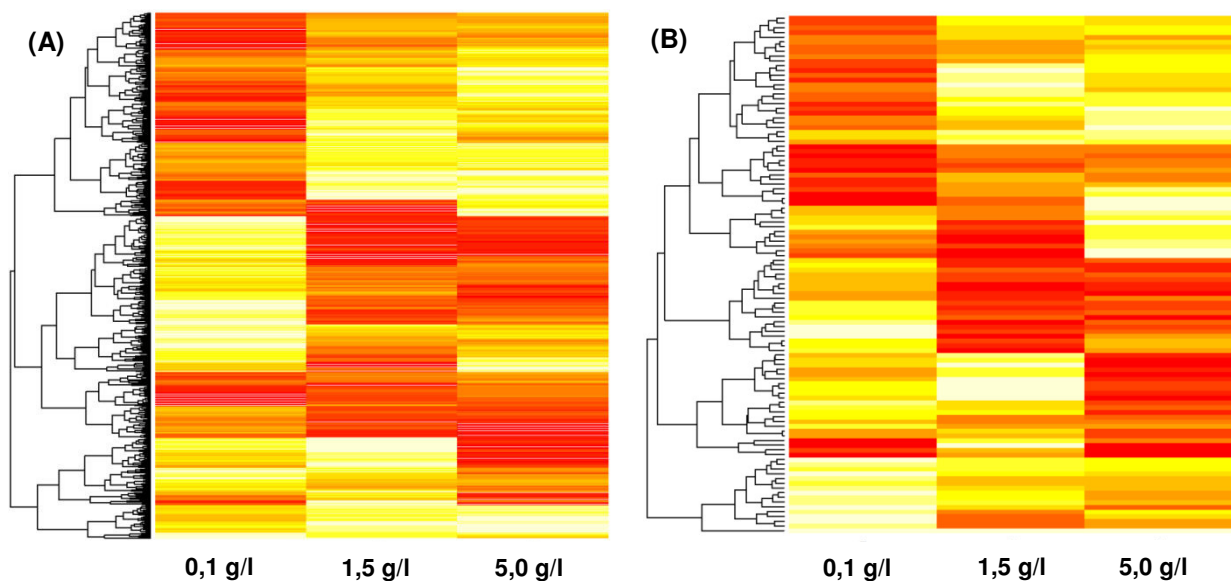


Figure 2: Heat map of three different samples from *Rhodobacter sphaeroides* strain 2.4.1, chromosome I (A) and II (B). Plotted are mean expression values of transcripts from cultures which were cultured with three different concentrations of glutamic acid (0.1, 1.5 and 5,0 g/l).

The gas production rate was observed to identify the influence on gene expression (see Figure 3 (A)). There seems to be an optimum at 1,5 [g/l] glutamic acid with almost 160 ml/hl⁻¹. At a glutamic acid concentration of 0,1 and 5,0 [g/l] significant lower H₂ evolution rates are observed (about 10 and 20 ml/hl⁻¹, respectively). The biomass increases nearly linear with accumulation of glutamic acid concentration (data kindly provided by the group of Jost Weber from the Technical University of Dresden). This leads to the conclusion that the activity of the bacteria is more important for hydrogen formation than the absolute cell mass. This could also be seen on transcriptional level.

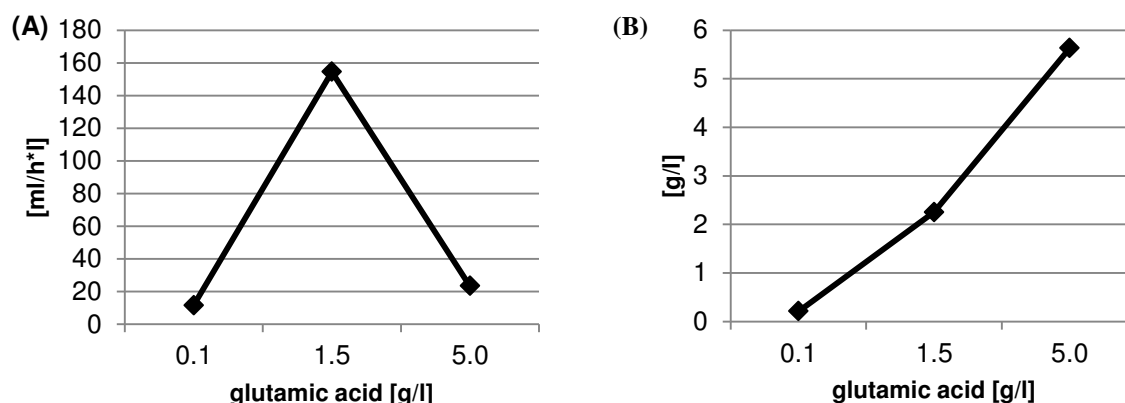


Figure 3: Hydrogen production rate (A) and bio dry mass (B) of *Rhodobacter sphaeroides* strain 2.4.1 during exponential growth under different concentrations of glutamic acid (The data were kindly provided by the group of Jost Weber from the Technical University of Dresden).

The *nitrogenase* enzyme mediates the production of hydrogen gas. We considered all structural *nitrogenase* genes, which are RS0110930 (*NifM*), RS0110935 (*NifE*), RS0110940 (*nitrogenase molybdenum-iron protein beta chain*), RS0110945 (*nitrogenase molybdenum-iron protein alpha chain*) and RS0110950 (*nitrogenase reductase*) on chromosome I. There are no genes located on chromosome II. In Figure 4, the mean expression values of all identified transcripts are shown. All of these show a correlating expression pattern to each other as well to the hydrogen evolution rate. It can be derived from the data that there is an increased expression of *nitrogenase* genes at a glutamic acid concentration of 1,5 [g/l]. This corresponds to the literature [2], where it was postulated that *nitrogenase* expression is increased during hydrogen production. It was also reported that there is no H₂ evolution with ammonium as nitrogen source [2]. This is confirmed by observations from the group of Jost Weber as well. Transcripts encoding for *nitrogenase* genes are higher expressed in H₂ producing cells. Furthermore, this enzyme is required for growth in hydrogen producing *Rhodobacter* cells.

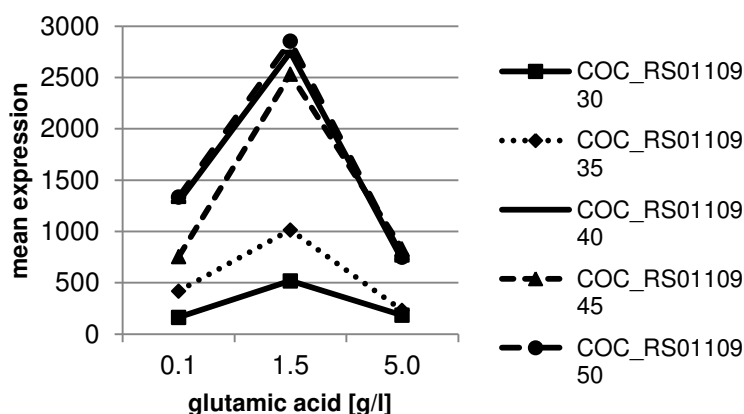


Figure 4: Mean expression of structure genes from the *nitrogenase* of *Rhodobacter sphaeroides* strain 2.4.1 on chromosome I during exponential growth under different concentrations of glutamic acid.

As mentioned before, the *hydrogenase* recycles excess hydrogen gas. Just like with the *nitrogenase*, we considered the expression of all structure genes of this enzyme (shown in Figure 5): RS0110735 (*hydrogenase* expression protein), RS0110740 (*hydrogenase*), RS0110750 (*hypB*), RS0110765 (*hypA*), RS0110780 (*hupF*), RS0110785 (*hypD*) and RS0110790 (*hupF*). The subunits of this enzyme show a similar expression behavior. With increasing content of glutamic acid up to 1,5 g/l, the expression of *hydrogenase* decreases almost to zero and rises slightly again until a glutamic acid content of 5,0 g/l. In the literature, it is postulated that there is a negative correlation between the expression of *hydrogenase* and hydrogen production, although a positive correlation was expected [5]. The expression level of this enzyme in this study is lowest when producing most hydrogen. This leads to the conclusion that only small rates of hydrogen will be recovered. That can be explained by removal of gas from the reactor.

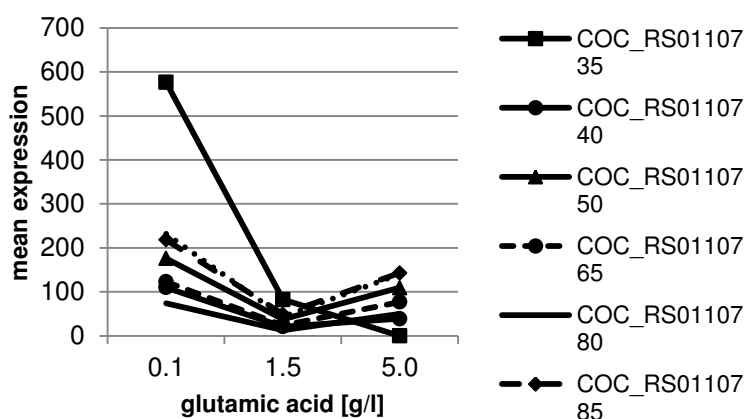


Figure 5: Mean expression of structure genes from the *hydrogenase* of *Rhodobacter sphaeroides* strain 2.4.1 on chromosome I during exponential growth under different concentrations of glutamic acid.

According to their expression pattern, enzymes of the CBB cycle are shown in Figure 6 (chromosome I A-C and chromosome II D). The structure genes of chromosome I are RS0103625 (*aldolase*), RS0105535 (*fructose-bisphosphate aldolase*), RS0105540 (*phosphoglycerate kinase*), RS0107965 (*transketolase*), RS0107980 (*glyceraldehyde-3-phosphate dehydrogenase*), RS0110265 (*fructose 1,6-bisphosphatase*), RS0114760 (*ribulose bisphosphate carboxylase small chain*), RS0114765 (*ribulose bisphosphate carboxylase*), RS0114770 (*fructose-bisphosphate aldolase*), RS0114775 (*phosphoribulokinase*) and RS0114780 (*fructose-1 6-bisphosphatase*). On chromosome II are coded the genes RS0116960 (*fructose-1 6-bisphosphatase*), RS0116965 (*phosphoribulokinase*), RS0116970 (*transketolase*), RS0116980 (*fructose-bisphosphate aldolase*) and RS0116985 (*ribulose bisphosphate carboxylase*).

The *ribulose bisphosphate carboxylase* (RS0114760, RS0114765 and RS0116985) is less expressed if there is a high hydrogen evolution rate [5]. This corresponds to our data. The reductive pentose phosphate pathway, especially the *ribulose 1,5-bisphosphate carboxylase-oxygenase* of the CBB cycle is used by photosynthetic bacteria for carbon dioxide reduction as well as building up cell mass. When hydrogen gas is produced and the nitrogen complex is derepressed, the redox potential of the *carboxylase-oxygenase* is balanced during growth. *Rhodobacter sphaeroides* is capable of photoheterotrophic growth in absence of a CBB pathway. This results in a high hydrogen gas evolution rate [2].

Furthermore, there are many transcripts of the CBB cycle that are negatively correlated, which applies to RS0114770, RS0114775, RS0116970 and RS0110116980. This was observed by Kontur *et al.* (2014) as well. In contrast, the fructose-1-6-bisphosphatase and the phosphoribulokinase (RS0114780, RS0116960, RS0116965) are described to have higher transcription levels during hydrogen production, which do not agree with our observations. This enzymes show a negatively expression pattern, too [5]. RS0110265, RS0105535, RS0105540, RS0107965 and RS0107980 are positively correlated to gas production, which was shown for the *fructose-1,6-bisphosphatase* (RS0110265) [5] as mentioned before. Only the *aldolase* (RS0103625) differs from all other expression patterns, which shows that this enzyme has a correlating expression with increasing glutamic acid concentration.

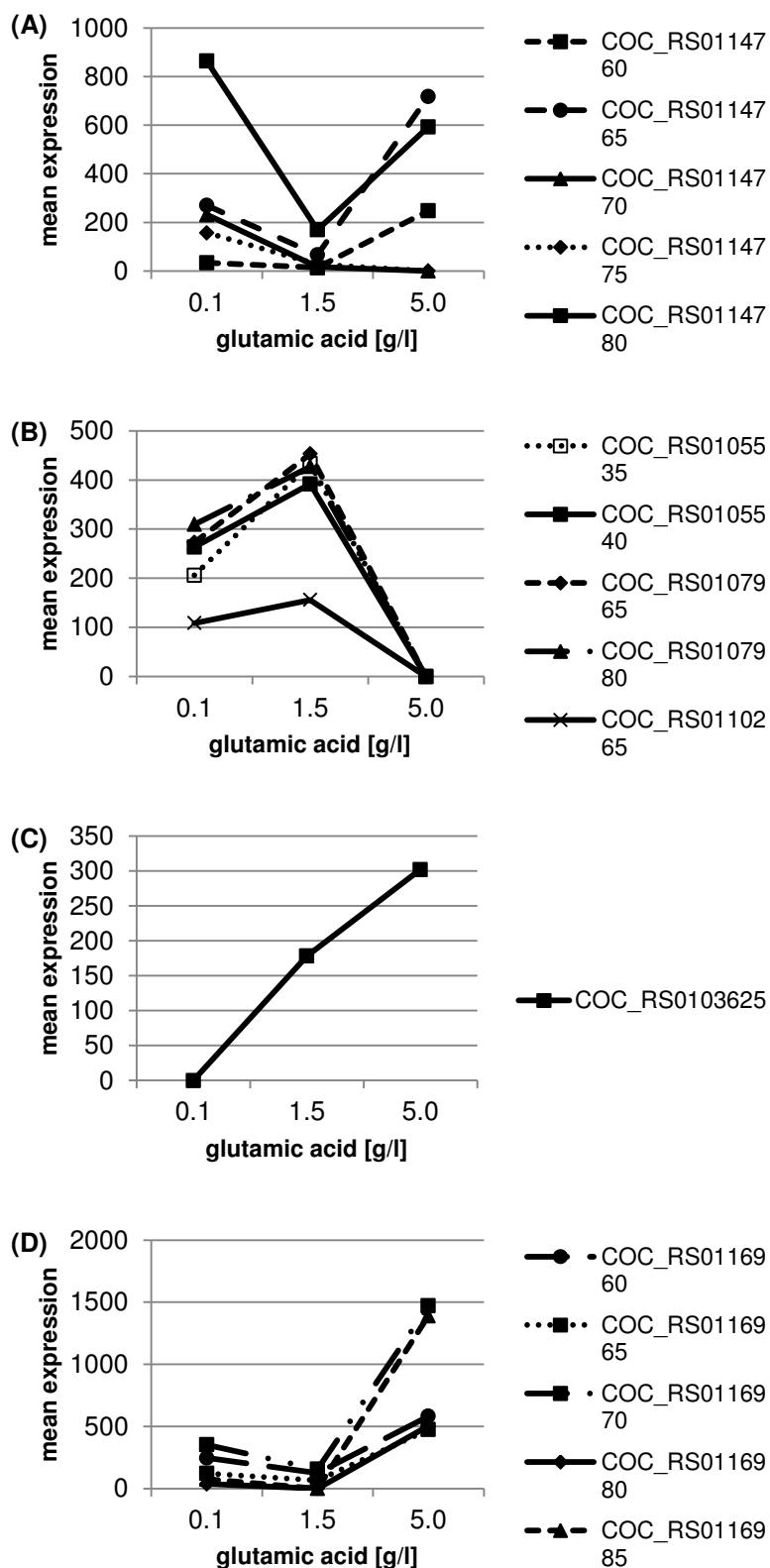


Figure 6: Mean expression of genes from Calvin-Benson-Bassham pathway of *Rhodobacter sphaeroides* strain 2.4.1 on chromosome I (A-C) and chromosome II (D) during exponential growth under different concentrations of glutamic acid. Transcripts are clustered by expression pattern.

For a general overview, all enzymes with known EC number were mapped onto the Biochemical Pathways Map (Boehringer) with CyanoMaps [14] (see Figure 7). Up-regulated genes are colored in red while down-regulated genes are colored in green. By this representation, metabolic pathways can be better understood. The complete map is available in the supplemental material, which is intended to give an overview of the regulation of all metabolic pathways in *Rhodobacter sphaeroides*. Figure 7 shows the regulation of enzymes which are involved in the CPP cycle. Many of them were up-regulated during exponential growth.

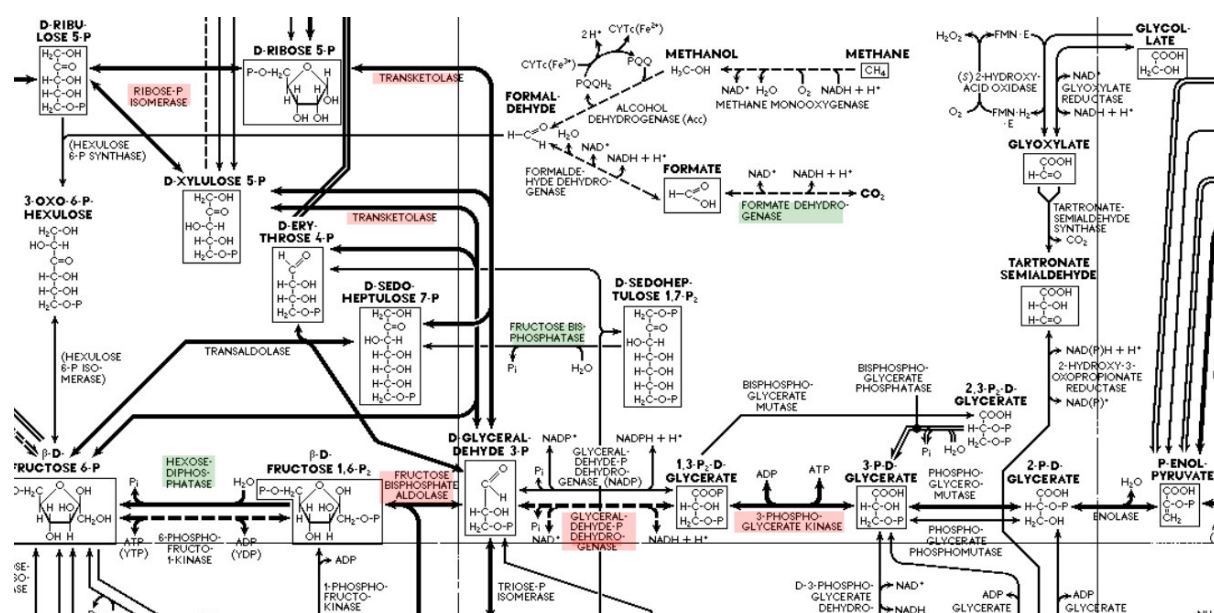


Figure 7: Detailed mapping of enzymes from CBB cycle from *Rhodobacter sphaeroides* strain 2.4.1 according to their regulation by CyanoMaps [14]. Up-regulated genes are colored in red, down-regulated ones in green.

For future studies, mutants of *Rhodobacter* should be taken into account. For example, the knock-out of *hydrogenase* and genes of CBB may increase hydrogen evolution rates. Furthermore, different nitrogen sources with several concentrations could be tested.

4 Conclusion

The transcriptome of *Rhodobacter sphaeroides* strain 2.4.1 was analyzed in context of biological hydrogen production. Cells were cultured in same medium which differed only in content of nitrogen source in terms of glutamic acid. For this reason the view was focused on enzymes which are involved in H₂ evolution pathways (*nitrogenase*, *hydrogenase* and *Calvin-Benson-Bassham* cycle). On all five plasmids there was no evidence for any gene expression. Chromosome I and II showed a highly differential gene expression in comparison of the three samples. The observed expression patterns are most closely corresponding to

further analysis of other working groups. It could be shown, that expression of genes involved in hydrogen production is highly influenced by glutamic acid concentration.

Acknowledgments

This work is performed in cooperation with the group of Jost Weber from the Technical University Dresden/Germany. For financial support we thank the State Ministry of Higher Education, Research and the Arts as well as the European Social Fund.

References

- [1] Franchi, Elisabetta; Tosi, Claudio; Scolia, Giuseppe; Penna, Gino Della.; Rodriguez, Francesco; Pedroni, Paola Maria (2004): Metabolically Engineered *Rhodobacter sphaeroides* RV strains for Improved Biohydrogen Photoproduction Combined with Disposal of Food Wastes. Mar. Biotechnol. 6, 552-565, DOI: 10.1007/s10126-004-1007-y.
- [2] Hallenbeck, Patrick C. (2010): Recent Advances in Phototrophic Prokaryotes. Springer Verlag. New York, p. 265-6.
- [3] Hellwig, Rico (2014): Prozessoptimierung am Chemostatsystem zur kontinuierlichen photobiologischen Wasserstoffproduktion unter Verwendung von *Rhodobacter sphaeroides* DSM 158. Masterthesis, University of Applied Sciences Mittweida.
- [4] Kim, Doong-Hoon; Lee, Ji-Hye; Kang, Seoktae; Hallenbeck, Patrick C.; Kim, Eui-Jin; Lee, Jeong K.; Kim, Mi-Sun (2014): Enhanced photo-fermentative H₂ production using *Rhodobacter sphaeroides* by ethanol addition and analysis of soluble microbial products. Biotechnology for Biofuels, 7:79.
- [5] Kontur, Wayne S.; Ziegelhoffer, Eva C.; Spero, Melanie A.; Imam, Saheed; Noguera, Daniel R.; Donohue, Timothy J. (2011): Pathways Involved in Reductant Distribution during Photobiological H₂ Production by *Rhodobacter sphaeroides*. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 2011, p. 7425–7429
- [6] Levin, David B.; Pitt, Lawrence; Love, Murray (2004): Biohydrogen production: prospects and limitations to practical application. International Journal of Hydrogen Energy 29, 173-85.
- [7] Lindberg, Pia (2003): Cyanobacterial Hydrogen Metabolism - Uptake Hydrogenase and Hydrogen Production by Nitrogenase in Filamentous Cyanobacteria. Acta Universitatis

Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 875. 52 pp. Uppsala.

Used software and version:

[8] FastQC, Version 0.11.2

[9] FastX-Toolkit, Version 0.0.13

[10] Bowtie2, Version 2.0.0-beta6

[11] Samtools, Version 1.0

[12] Artemis

[13] R

[14] CyanoMaps (not yet published)

4 Ausblick

Das langfristige Ziel dieses Projektes ist die Herstellung von Wasserstoff mit Hilfe der Co-Kultivierung von Grünalgen und *Rhodobacter sphaeroides* im industriellen Maßstab. Diese Methode der Energiegewinnung ist besonders umweltschonend, da keine schädlichen Nebenprodukte entstehen. In Abbildung 5 ist der schematische Aufbau eines solchen Versuchs dargestellt. Energie wird von Algen als auch von *Rhodobacter sphaeroides* aus Licht (im optimalen Fall aus Sonnenlicht mittels Photobioreaktoren), gewonnen. Mit Hilfe von Wasser und Kohlenstoffdioxid produzieren die Algen Kohlenstoffäquivalente, welche dem Bioreaktor mit *Rhodobacter sphaeroides* zugeführt wird. Diese werden anschließend durch die Nitrogenase in Wasserstoff und Kohlendioxid umgewandelt. Das Kohlendioxid wird im Anschluss wieder der Algenfermentation zugeführt. Der entstandene Wasserstoff kann abgeleitet und als Energiequelle, beispielsweise für den Einsatz in Brennstoffzellen, verwendet werden.

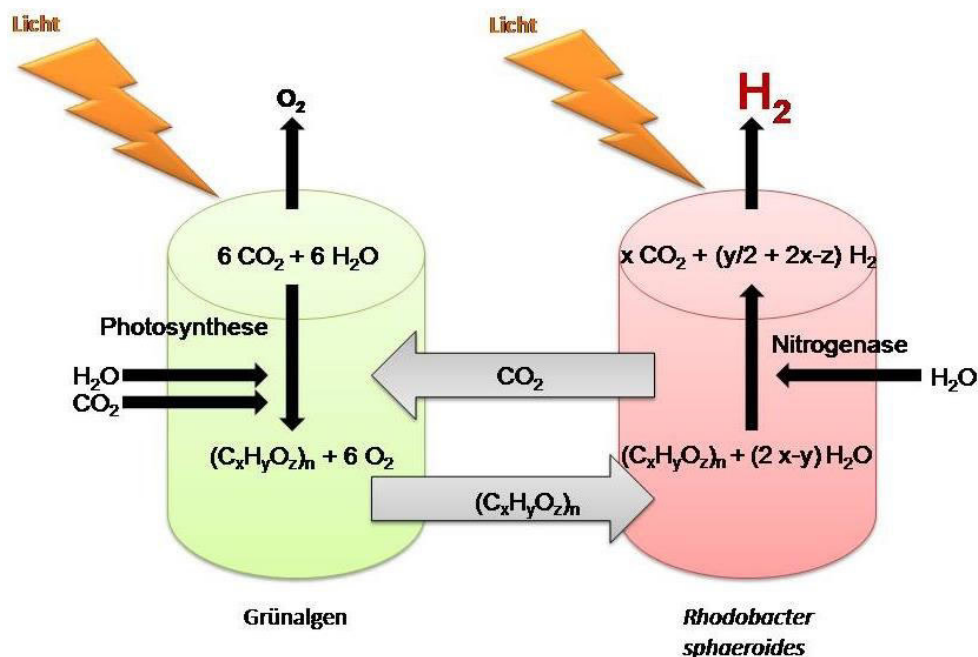


Abbildung 5: Schema des Prozesses der Wasserstoffgewinnung aus *Rhodobacter sphaeroides* mit Hilfe von Grünalgen. *R. sphaeroides* nutzt dabei Kohlenhydrat-Äquivalente aus dem Metabolismus der Grünalgen. Wasserstoff wird abgeleitet und Kohlenstoffdioxid aus der Bakterienfermentation wieder den Grünalgen zugeführt.

Für die zukünftige Verwendung von Wasserstoff als alternativen Energieträger, können die resultierenden Wasserstoffraten durch Prozessoptimierungen weiter erhöht werden. Beispielsweise sollten Mutationsstudien mit *Rhodobacter sphaeroides* durchgeführt werden. Resultierend aus den Ergebnissen dieser Arbeit zählen dazu der Knock-out der

Hydrogenase, um die Wasserstoffrückführung zu unterbinden, sowie die Abschaltung der Gene des CBB-Zyklus, da das Bakterium dann in der Lage ist photoheterotroph zu wachsen. Daraus resultiert eine erhöhte Wasserstoffproduktionsrate. Weiterhin sollten auch andere Substrate mit unterschiedlichen Konzentrationen auf Eignung für die H₂-Entstehung getestet werden.

Weiterführende experimentelle Untersuchungen mit anderen Stämmen sollten zeigen, inwieweit sich diese in ihrem Vermögen Wasserstoff zu bilden, unterscheiden. Dies könnte weitere Anhaltspunkte für Ursachen veränderlicher Wasserstoffraten geben. Speziell sollten die Stoffwechselwege der Stämme untersucht werden, welche für die Bildung von H₂ verantwortlich oder an diesen beteiligt sind. Erkenntnisse könnten anschließend auf den Stamm 2.4.1 übertragen werden.

5 Zusammenfassung

Das langfristige Ziel dieser Arbeit ist die Optimierung des Fermentationsprozesses von *Rhodobacter sphaeroides* Stamm 2.4.1, um maximale Wasserstoffproduktionsraten zu erreichen. Die Co-Kultivierung von Grünalgen und dem untersuchten *Rhodobacter* Stamm ist umweltschonend. Dabei entstehen keine schädlichen Nebenprodukte. Diese Art der Wasserstoffproduktion könnte zukünftig eine Alternative zu fossilen Energieträgern sein. In dieser Arbeit wurden erste Analysen auf Genom- und Transkriptomebene durchgeführt, um Einblicke in Regulationswege der Wasserstoffentstehung zu gewinnen.

Ziel einer Genomsequenzierung ist die Identifizierung bislang unbekannter Gene, sowie die Erkennung von Veränderungen im Genom innerhalb eines Organismus. So genannte *single nucleotide polymorphisms* (kurz SNPs) sind verantwortlich für die Diversität der Organismen (Brown, 1999). In dieser Arbeit wurde das gesamte Genom von *Rhodobacter sphaeroides* Stamm 2.4.1 re-sequenziert. Ziel war die Suche nach Ursachen für variierende Wasserstoffraten dieses Stammes aus unterschiedlichen Laboren. Oft unterscheiden sich experimentelle Ergebnisse gleicher Stämme aufgrund unterschiedlicher Entwicklungen in. Es konnten 222 SNPs identifiziert werden. 71 davon befanden sich innerhalb von Genen und führten zu Veränderungen in codierenden Bereichen des Genoms. Diese waren entweder neutral oder führten zu Aminosäureaustauschen. Auf Chromosom I wurden insgesamt 18 Enzyme identifiziert, deren protein-codierende Sequenzen durch SNPs verändert waren. Keines der veränderten Enzyme ist direkt in den Prozess der Wasserstoffproduktion involviert. Ein indirekter Einfluss durch Proteine, bei welchen strukturelle Veränderungen zu veränderter oder eingeschränkter Funktion geführt haben, kann nicht ausgeschlossen werden. Diese, sowie SNPs in intergenischen Bereichen, wurden in dieser Arbeit nicht weiter untersucht. Veränderungen in Intronbereichen können einen Einfluss auf das Transkriptionslevel von Genen haben, wenn sie in Promotorbereichen lokalisiert sind. Auch SNPs innerhalb hypothetischer Proteine könnten einen Einfluss auf Wasserstoffproduktionsraten haben, wenn sie in beteiligte Stoffwechselwege involviert sind. Die Integrierung genetischen Materials, horizontaler DNA Transfer oder Neuordnungen genetischer Loci führen zu unterschiedlichen Entwicklungen von Chromosomen oder ganzer Genome (Choudhary et al., 2007). Solche Veränderungen entstehen durch Anpassung an unterschiedliche Lebensräume und Umweltbedingungen und führt zu Abwandlungen von Stoffwechselwegen.

Eine Analyse zum Vergleich der bekannten *Rhodobacter* Stämme mittels eines multiplen Sequenzalignments zeigte, dass es innerhalb des Referenzstamms 2.4.1 und dem re-sequenzierten Stamm, keine Neuorganisation von Genen innerhalb des Genoms kam. Daher gibt auch dieses Ergebnis keinen Anhalt für die Ursache veränderlicher

Wasserstoffentsehungsraten. Der Genomvergleich mit anderen *Rhodobacter sphaeroides* Stämmen zeigt zahlreiche Veränderungen und Umorganisationen innerhalb der Genome.

Daher wurde im zweiten Schritt auch das Transkriptom von *Rhodobacter sphaeroides* Stamm 2.4.1 sequenziert und im Hinblick auf die biologische Wasserstoffproduktion analysiert. Drei unterschiedliche Fermentationsansätze wurden untersucht und verglichen. Diese unterschieden sich nur im Gehalt an Glutaminsäure, welche *Rhodobacter* als Substrat diente. Auch hier wurde sich bei der Auswertung der Daten auf Proteine (beziehungsweise Enzyme) beschränkt, welche direkt in den Prozess der Wasserstoffentstehung involviert sind. Dies waren Strukturgene der Nitrogenase, Hydrogenase und des Calvin-Benson-Bassham Stoffwechselweges. Auf keinem der Plasmide konnte Genexpression nachgewiesen werden, sodass die Auswertung auf Chromosom I und II beschränkt werden konnte. Auf beiden Chromosomen konnte eine stark differentielle Expression von Transkripten im Vergleich der drei Ansätze nachgewiesen werden. Auch Gene der Nitrogenase, Hydrogenase sowie Enzyme des CBB-Stoffwechselweges zeigen bei unterschiedlichen Glutaminsäurekonzentrationen stark differenzierende Expressionswerte. Dies lässt auf einen starken Einfluss des verwendeten Substrats und dessen Konzentration während der Kultivierung schließen. Das Optimum der Wasserstoffproduktion lag bei 160 ml Gas pro Stunde und Liter Medium bei einer Glutaminsäurekonzentration von 1,5 g/l. Daher sollte diese Substratkonzentration in zukünftigen Versuchen gewählt werden.

Das Enzym Nitrogenase vermittelt die Wasserstoffproduktion unter Verwertung von Licht (Lindberg, 2003) und organischen Säuren (Levin et al., 2004) als Substrat. Dabei ist Wasserstoff ein Nebenprodukt der Stickstofffixierung des Enzyms (Kontur et al., 2011). Neben der Nitrogenase ist auch die Hydrogenase in den H_2 Entstehungsprozess involviert. Dieses Enzym führt überschüssigen Wasserstoff zurück, während Gene des CBB-Stoffwechselweges Kohlendioxid reduzieren und Zellmasse bilden (siehe Abbildung 1) (Franchi et al., 2004). Die Nitrogenaseexpression korreliert dabei mit der Wasserstoffproduktion. Das heißt je höher die Gasrate, umso mehr Enzym wird exprimiert. Bei der Hydrogenaseexpression ist ein umgekehrter Effekt zu beobachten. Allerdings wurde eine erhöhte Rate der Hydrogenaseexpression erwartet, da die Hydrogenase die Funktion der Wasserstoffrückführung ausübt. Das in dieser Arbeit beobachtete Ergebnis wurde allerdings auch von Kontur et al. (2011) beschrieben (Kontur et al., 2011). Dies kann auf die Ableitung des Wasserstoffs vom Reaktor zurückgeführt werden. Zusammenfassend kann gesagt werden, dass das Expressionsverhalten der untersuchten Gene weitestgehend mit Ergebnissen anderer Arbeitsgruppen korreliert (siehe Kontur et al., 2011).

Literaturverzeichnis

Anders, Simon; Huber, Wolfgang (2010): Differential expression analysis for sequence count data. *Gen Biol*, 11:R106.

Andrews, Simon (17.09.2014): FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Azam, Sarwar; Rathore, Abhishek; Shah, Trushar M.; Telluri, Mohan; Amindala, Bhanu Prakash; Rperao, Pradeep; Katta, Mohan A. V. S. K.; Varshney, Rajeev K. (2014): An Integrated SNP Mining and Utilization (ISMU) Pipeline for Next Generation Sequencing Data. *PLOS ONE*, Vol. 9-e101754.

Brown, Terence A. (1999): *Genomes*. John Wiley & Sons, Inc, New York.

Caver, Tim; Velarde, Giles; Berriman, Matthew; Parkhill, Julian; McQuillan, Jaqueline A. (2010): Artemis and ACT—browsing genomes and visualisation of next generation data. Pathogen Genomics Group, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UK.

Choudhary, M.; Zanhua, Xie; Fu, Y. X.; Kaplan, S. (2007): Genome Analyses of Three Strains of *Rhodobacter sphaeroides*: Evidence of Rapid Evolution of Chromosome II. *Journal of Bacteriology*. p. 1914-1921.

Cock, Peter J. A.; Fields, Christopher J.; Goto, Naohisa; Heuer, Michael L.; Rice, Peter M.: (2010): The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, Vol. 38, No. 6 1767–71.

Das, Debrata; Veziroğlu (2001): Hydrogen production by biological processes: a survey of literature. *Int Jour of Hydrogen Energy* 26, 13-28.

Ferraz Júnior, Antônio Djalma; Zaiat, Marcelo; Gupta, Medhavi; Elbeshbishy, Elsayed; Hafez, Hisham; Nakhla, George (2014): Impact of organic loading rate on biohydrogen production in an up-flow anaerobic packed bed reactor (UANPBR). *Bioresource Technology* 164: 371-9.

Franchi, Elisabetta; Tosi, Claudio; Scolia, Giuseppe; Penna, Gino Della.; Rodriguez, Francesco; Pedroni, Paola Maria (2004): Metabolically Engineered *Rhodobacter sphaeroides* RV strains for Improved Biohydrogen Photoproduction Combined with Disposal of Food Wastes. *Mar. Biotechnol.* 6, 552-565, DOI: 10.1007/s10126-004-1007-y.

Giotta, Livia; Agostiano, Angela; Italiano, Francesca; Milano, Francesco; Trotta, Massimo (2006): Heavy metal ion influence on the photosynthetic growth of *Rhodobacter sphaeroides*. *j.chemosphere*, Vol. 62, Issue 9, 1490–99.

Illumina, Inc. (2013): *An Introduction to Next-Generation Sequencing Technology* [online]. Pub No. 770-2012-008.

Kim, Dong-Hoon Kim; Lee, Ji-Hye; Kang, Seoktae Kang; Hallenbeck, Patrick C.; Kim, Eui-Jin; Lee, Jeong K.; Kim, Mi-Sun Kim (2014): Enhanced photo-fermentative H₂ production using *Rhodobacter sphaeroides* by ethanol addition and analysis of soluble microbial products. *Biotechnology for Biofuels*, 7:79.

Kontur, Wayne S.; Schackwitz, Wendy S.; Ivanova, Natalia; Martin, Joel; LaButti, Kurt; Deshpande, Shweta; Tice, Hope N.; Pennacchio, Christa; Sodergren, Erica; Weinstock, George M.; Noguera, Daniel R.; Donohue, Timothy J. (2012): Revised Sequence and Annotation of the *Rhodobacter sphaeroides* 2.4.1 Genome. *J Bacteriol.*, 194(24): 7016–7017.

Kontur, Wayne S.; Ziegelhoffer, Eva C.; Spero, Melanie A.; Imam, Saheed; Noguera, Daniel R.; Donohue, Timothy J. (2011): Pathways Involved in Reductant Distribution during Photobiological H₂ Production by *Rhodobacter sphaeroides*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, Oct. 2011, p. 7425–7429.

Levin, David B.; Pitt, Lawrence; Love, Murray (2004): Biohydrogen production: prospects and limitations to practical application. *International Journal of Hydrogen Energy* 29, 173-85.

Li, Heng; Handsaker, Bob; Wysoker, Alec; Fennell, Tim; Ruan, Jue; Homer, Nils; Marth, Gabor; Abecasis, Goncalo; Durbin, Richard (2009): The Sequence Alignment/Map format and SAMtools. *Vol. 25, No. 16*, 2078–79.

Lindberg, Pia (2003): Cyanobacterial Hydrogen Metabolism - Uptake Hydrogenase and Hydrogen Production by Nitrogenase in Filamentous Cyanobacteria. *Acta Universitatis*

Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 875. 52 pp. Uppsala.

Liu, Lin; Li, Yinhu; Li, Siliang; Hu, Ni; He, Yimin; Pong, Ray; Lin, Danni; Lu, Lihua; Law, Maggie (2012): Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*. Article ID 251364.

Metz, Sebastian (2010): Blaulichtabhängige Genregulation in *Rhodobacter sphaeroides*. Untersuchungen zur physiologischen Funktion der Blaulichthotorezeptoren AppA, CryB und LOV. Inaugural - Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften. Justus - Liebig - Universität Gießen.

Montagud, Arnau; Gamermann, Daniel; Fernández de Córdoba, Pedro; Urchueguia, Javier F. (2013): *Synechocystis* sp. PCC6803 metabolic models for the enhanced production of hydrogen. doi:10.3109/07388551.2013.829799.

Porter, Steven L.; Wilkinson, David A.; Byles, Elaine D.; Wadhams, George H.; Taylor, Stephen; Saunders, Nigel J.; Armitage, Judith P. (2011): Genome Sequence of *Rhodobacter sphaeroides* Strain WS8N. *J Bacteriol.*, 193(15): 4027–4028.

Schmieder, Robert; Edwards, Roberts (2011): Quality control and preprocessing of metagenomic datasets. *Bioinformatics Application Notes*. Vol. 27 No. 6, 863–4.

URL-1 (05.03.2014): *Rhodobacter sphaeroides* 2.4.1, ATCC BAA-808. URL: <http://genome.jgi-psf.org/rhosp/rhosp.gif>.

Wang, Z.; Gerstein, M.; Snyder, M. (2009): RNA-Seq: a revolutionary tool for Transcriptomics. *NATURE REVIEWS|GENETICS*, Vol. 10, 57-63.

Wink, Michael (2011): *Molekulare Biotechnologie. Konzepte, Methoden und Anwendungen*. 2., aktualisierte Auflage. Wiley-VHC Verlag, Weinheim.

Wünschiers, R. (2013): *Computational Biology. A Practical Introduction to BioData Processing and Analysis with Linux, MySQL and R*. Second Edition: Springer Verlag, Heidelberg.

Yang, M.K., Lin, Y.C., Shen, C.H. (2006): Identification of two gene loci involved in poly-beta-hydroxybutyrate production in *Rhodobacter sphaeroides* FJ1. *J Microbiol Immunol Infect.* 39(1):18-27.

Zhou, Shiguo; Kvikstad, Erika; Kile, Andrew (2003): Whole-Genome Shotgun Optical Mapping of *Rhodobacter sphaeroides* strain 2.4.1 and Its Use for Whole-Genome Shotgun Sequence Assembly. *Genome Res.* 13: 2142-2151.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, den 30. November 2014

A handwritten signature in black ink, reading "Nadine Wappler". The signature is written in a cursive style with a large, stylized 'W'.

Nadine Wappler